



TESIS SS14 2501

***RARE EVENT WEIGHTED LOGISTIC REGRESSION
UNTUK KLASIFIKASI *IMBALANCED DATA****
(Studi Kasus: Klasifikasi Desa Tertinggal di Provinsi Jawa Timur)

DIAN EKA APRIANA SULASIH
NRP 1314 201 714

DOSEN PEMBIMBING
Santi Wulan Purnami, M.Si., Ph.D.
Santi Puteri Rahayu, M.Si., Ph.D.

PROGRAM MAGISTER
JURUSAN STATISTIKA
FAKULTAS MATEMATIKA DAN ILMU PENGETAHUAN ALAM
INSTITUT TEKNOLOGI SEPULUH NOPEMBER
SURABAYA
2016



THESIS SS14 2501

**RARE EVENT WEIGHTED LOGISTIC REGRESSION
FOR CLASSIFICATION OF IMBALANCED DATA
(Case Study: The Classification of Underdeveloped Rural
In East Java Province)**

**DIAN EKA APRIANA SULASIH
NRP 1314 201 714**

**SUPERVISOR
Santi Wulan Purnami, M.Si., Ph.D.
Santi Puteri Rahayu, M.Si., Ph.D.**

**PROGRAM OF MAGISTER
STATISTICS DEPARTMENT
FACULTY OF MATHEMATICS AND NATURAL SCIENCE
SEPULUH NOPEMBER INSTITUTE OF TECHNOLOGY
SURABAYA
2016**

RARE EVENT WEIGHTED LOGISTIC REGRESSION
UNTUK KLASIFIKASI *IMBALANCED DATA*
(Studi Kasus: Klasifikasi Desa Tertinggal di Provinsi Jawa Timur)

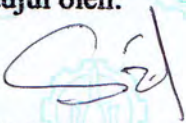
Tesis disusun untuk memenuhi salah satu syarat memperoleh gelar
Magister Sains (M. Si.)

di
Institut Teknologi Sepuluh Nopember
Oleh :

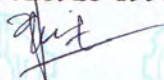
DIAN EKA APRIANA SULASIH
NRP. 1314 201 714

Tanggal Ujian : 26 Januari 2016
Periode Wisuda : Maret 2016


Disetujui oleh:


1. Santi Wulan Purnami, M.Si., Ph.D.
NIP. 19720923 199803 2 001

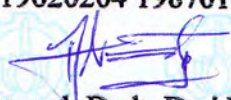
(Pembimbing I)


2. Santi Puteri Rahayu, M.Si., Ph.D.
NIP. 19750115 199903 2 003

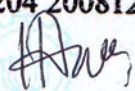
(Pembimbing II)


3. Dr. Purnadi, M.Sc.
NIP. 19620204 198701 1 001

(Penguji)


4. Dr. rer. pol. Dedy Dwi Prastyo, M.Si.
NIP. 19831204 200812 1 002

(Penguji)


5. Dr. Margaretha Ari Anggorowati S.Kom, M.T.
NIP. 19720222 199803 2 002

(Penguji)



Direktur Program Pascasarjana,


Prof. Ir. Djauhar Manfaat, M. Sc., Ph. D.
NIP. 19601202 198701 1 001

RARE EVENT WEIGHTED LOGISTIC REGRESSION
UNTUK KLASIFIKASI *IMBALANCED DATA*
(Studi Kasus: Klasifikasi Desa Tertinggal
di Provinsi Jawa Timur)

Nama : Dian Eka Apriana Sulasih
NRP : 1314201714
Pembimbing : Santi Wulan Purnami, M.Si., Ph.D.
Santi Puteri Rahayu, M.Si., Ph.D.

ABSTRAK

Salah satu permasalahan dalam klasifikasi data adalah komposisi data yang tidak seimbang (*imbalanced data*). Pada klasifikasi *imbalanced data*, *classifier* cenderung memprediksi kelas yang memiliki komposisi data lebih besar sehingga didapatkan akurasi prediksi yang baik terhadap kelas data *training* yang banyak (kelas mayoritas) dan akurasi prediksi yang buruk untuk kelas data *training* yang sedikit (kelas minoritas). Oleh karena itu, diperlukan metode yang tepat untuk melakukan klasifikasi pada *imbalanced data*. *Rare Event Weighted Logistic Regression (RE-WLR)* adalah metode klasifikasi *imbalanced data* untuk data berukuran besar dan *rare event*. *RE-WLR* dikembangkan dari *Truncated Regularized Iteratively Re-weighted Least Square (TR-IRLS)* dengan *rare event correction* pada Regresi Logistik. Penelitian ini bertujuan untuk mengkaji dan menerapkan *RE-WLR* untuk klasifikasi *imbalanced data* dengan studi kasus klasifikasi desa tertinggal di Provinsi Jawa Timur tahun 2014, serta untuk membandingkan tingkat ketepatan klasifikasi antara metode *RE-WLR* dan *TR-IRLS* pada kasus tersebut. Hasil penelitian menunjukkan bahwa secara deskriptif *RE-WLR* memberikan kinerja klasifikasi yang lebih baik dibandingkan *TR-IRLS*, namun dengan perbedaan yang tidak signifikan. Rata-rata nilai *sensitivity RE-WLR* juga lebih tinggi daripada *TR-IRLS*. Hal ini menunjukkan bahwa *RE-WLR* bisa memprediksi kelas minoritas (*rare event*) atau desa tertinggal dengan lebih baik dibandingkan *TR-IRLS*.

Kata kunci : Desa Tertinggal, *Imbalanced Data*, Klasifikasi, *RE-WLR*, *TR-IRLS*

**RARE EVENT WEIGHTED LOGISTIC REGRESSION
FOR CLASSIFICATION OF IMBALANCED DATA
(Case Study: The Classification of Underdeveloped Rural
In East Java Province)**

Name : Dian Eka Apriana Sulasih
NRP : 1314201714
Supervisor : Santi Wulan Purnami, M.Si., Ph.D.
Santi Puteri Rahayu, M.Si., Ph.D.

ABSTRACT

One of the problems in data classification is the composition of the data that is out of balance (imbalanced data). In the classification of imbalanced data, most of the classifier are biased towards the major class and have very poor classification rates on minor class. Rare Event Weighted Logistic Regression (RE-WLR) is a method of classification applied to large imbalanced data and rare event. RE-WLR is developed from Truncated Regularized Iteratively Re-weighted Least Squares (TR-IRLS) with rare event correction to Logistic Regression. This study aims to assess and apply the RE-WLR to the classification of imbalanced data with study case classification of underdeveloped rural in East Java Province in 2014, and to compare the accuracy between RE-WLR method and TR-IRLS in that case. The results shows that RE-WLR provides better classification performance than TR-IRLS, but the difference is not significant. The average value of RE-WLR's sensitivity is also higher than TR-IRLS. This shows that the RE-WLR could predict the minority class (rare event) or underdeveloped rural better than TR-IRLS.

Keywords : Classification, Imbalanced Data, RE-WLR, TR-IRLS, Underdeveloped Rural

KATA PENGANTAR

Alhamdulillah Robbil ‘Aalamiin, puji syukur atas kehadiran Allah SWT, atas segala karunia nikmat, taufik dan hidayah-Nya kepada penulis sehingga tesis yang berjudul “***RARE EVENT WEIGHTED LOGISTIC REGRESSION UNTUK KLASIFIKASI IMBALANCED DATA (Studi Kasus: Klasifikasi Desa Tertinggal di Provinsi Jawa Timur)***” dapat terselesaikan. Dalam penyusunan tesis ini, penulis memperoleh banyak bantuan dari berbagai pihak, baik secara langsung maupun tidak langsung. Untuk itu pada kesempatan ini penulis ingin menyampaikan penghargaan dan ucapan terima kasih yang sebesar-besarnya kepada:

1. Badan Pusat Statistik (BPS) yang telah memberi kesempatan serta beasiswa kepada penulis untuk melanjutkan studi program S2 di ITS.
2. Ibu Santi Wulan Purnami, M.Si, Ph. D. dan Ibu Santi Puteri Rahayu, M.Si., Ph.D. selaku dosen pembimbing yang telah banyak meluangkan waktu serta dengan penuh kesabaran dan keikhlasan dalam memberikan bimbingan, saran dan masukan serta motivasi kepada penulis dalam penyusunan tesis ini.
3. Bapak Dr. Suhartono, M.Sc selaku Ketua Jurusan Statistika FMIPA ITS Surabaya yang telah memberikan fasilitas dan pelayanan yang baik selama penulis menyelesaikan pendidikan S2 di kampus tercinta ini.
4. Bapak Dr. Purhadi, M.Sc., Dr. rer. pol. Dedy Dwi Prastyo, M.Si. dan Ibu Dr Margaretha Ari Anggorowati S.Kom, M.T. selaku dosen penguji yang telah memberikan koreksi, saran, dan masukan dalam penyusunan tesis ini.
5. Bapak dan Ibu dosen Statistika atas pembekalan ilmu dan pengalamannya selama penulis menempuh pendidikan di Program Studi Magister Jurusan Statistika FMIPA ITS Surabaya.
6. Kedua orangtua tercinta, Bapak dan Ibu yang telah membesarkan, mendidik dan mendoakan dengan penuh keikhlasan dan kasih sayang, Adik tersayang serta semua keluarga yang telah memberikan dukungan, motivasi, semangat dan doanya.

7. Teman-teman kos ARH48: Santi, Yani, Mpih, Mbak Nike, Mbak Widi, dan Yanti, untuk segala bantuan, kerjasama, keceriaan, dan kebersamaannya selama ini. Semoga dapat berjumpa lagi di lain kesempatan.
8. Teman-teman BPS angkatan 8, Vivin, Mbak Afni, Mbak Nita, Maul, Mas Arip, Mas Mur, Mas Ali, Aan, Fatih, Mas Zablin, Mas Duto, Mas Henri, Bang Rory, terima kasih atas segala bantuan, dukungan, kebersamaan dan kekompakannya selama menjalani pendidikan di ITS, senang bisa bertemu dan mengenal teman-teman semua, semoga dapat berjumpa lagi di lain kesempatan.
9. Teman-teman BPS angkatan 9, teman-teman reguler angkatan 2014, Pak Irul, dan Bagian Administarsi Jurusan Statistika FMIPA ITS beserta semua pihak yang tidak bisa disebutkan satu per satu terima kasih atas bantuannya.
10. Semua teman dan kerabat lain yang telah mendukung, mendoakan, dan memotivasi penulis.

Akhir kata, semoga segala kebaikan yang telah diberikan kepada penulis, mendapatkan pahala dari Allah SWT. Penulis menyadari bahwa tesis ini masih jauh dari kesempurnaan, untuk itu kritik dan saran yang bersifat membangun penulis harapkan demi perbaikan pada masa yang akan datang. Penulis berharap semoga tesis ini dapat memberikan manfaat kepada pembaca serta menambah wawasan keilmuan untuk penulis sendiri. Aamiin Yaa Robbal 'Aalamiin.

Surabaya, Pebruari 2016

Penulis

DAFTAR ISI

LEMBAR PENGESAHAN.....	v
ABSTRAK	vii
ABSTRACT.....	ix
KATA PENGANTAR.....	xi
DAFTAR ISI.....	xiii
DAFTAR GAMBAR.....	xv
DAFTAR TABEL.....	xvii
DAFTAR LAMPIRAN.....	ix
BAB 1 PENDAHULUAN.....	1
1.1 Latar Belakang.....	1
1.2 Rumusan Masalah	5
1.3 Tujuan Penelitian	5
1.4 Manfaat Penelitian	6
1.5 Batasan Masalah Penelitian.....	6
BAB 2 TINJAUAN PUSTAKA.....	7
2.1 <i>Logistic Regession (LR)</i>	7
2.2 <i>Regularized Logistic Regression (LR)</i>	8
2.3 <i>Truncated Regularized Iteratively Re-weighted Least Square...</i>	8
2.4 <i>Rare Event Weighted Logistic Regression (RE-WLR)</i>	11
2.5 Evaluasi Kinerja.....	16
2.6 Uji Wilcoxon.....	18
2.7 Gambaran Umum Provinsi Jawa Timur.....	18
2.8 Desa Tertinggal.....	22
BAB 3 METODOLOGI PENELITIAN.....	27
3.1 Penaksiran Parameter Rare Event Weighted Logistic Regression	27
3.2 Penerapan Rare Event Weighted Logistic Regression.....	28
3.2.1 Sumber Data.....	28
3.2.2 Variabel Penelitian.....	29
3.2.3 Langkah-langkah Penerapan Rare Event Weighted	

Logistic Regression	68
BAB 4 HASIL DAN PEMBAHASAN.....	71
4.1 Penaksiran Parameter Model <i>RE-WLR</i>	71
4.2 Klasifikasi Desa Tertinggal di Provinsi Jawa Timur.....	80
4.2.1 Gambaran Umum Provinsi Jawa Timur Menurut Variabel Penelitian	80
4.2.2 Klasifikasi Desa Tertinggal di Provinsi Jawa Timur dengan <i>RE-WLR</i>	82
4.3 Perbandingan Ketepatan Klasifikasi Desa Tertinggal antara Model <i>TR-IRLS</i> dan <i>RE-WLR</i>	89
BAB 5 KESIMPULAN DAN SARAN.....	93
5.1 Kesimpulan.....	93
5.2 Saran.....	94
DAFTAR PUSTAKA.....	95
LAMPIRAN.....	101

DAFTAR GAMBAR

Gambar 2.1 Kurva ROC.....	17
Gambar 2.2 Peta Wilayah Administrasi Provinsi Jawa Timur.....	19
Gambar 4.1 Jumlah Penduduk Miskin Menurut Provinsi di Indonesia Tahun 2014.....	80
Gambar 4.2 Persentase Penduduk Miskin Menurut Lokasi Tempat Tinggal di Provinsi Jawa Timur Tahun 2014.....	81
Gambar 4.3 Persentase Desa Tertinggal dan desa Tidak Tertinggal di Provinsi Jawa Timur Tahun 2014.....	81
Gambar 4.4 Persentase Desa Tertinggal dan Desa Tidak Tertinggal Menurut Kabupaten/Kota di Provinsi Jawa Timur Tahun 2014.....	81

DAFTAR TABEL

Tabel 2.1 <i>Confusion Matrix</i>	16
Tabel 2.2 Jumlah, Laju Pertumbuhan dan Kepadatan Penduduk menurut Kabupaten/Kota di Jawa Timur Tahun 2014.....	21
Tabel 3.1 Variabel dalam Penelitian.....	29
Tabel 3.2 Struktur Data dalam Penelitian.....	30
Tabel 4.1 Data Jumlah Kelas Positif dan Negatif pada Data <i>Training</i> dan Data <i>Testing</i>	83
Tabel 4.2 Rata-rata Kinerja Klasifikasi <i>RE-WLR</i> untuk Klasifikasi Desa Tertinggal di Jawa Timur.....	84
Tabel 4.3 Perbandingan Kinerja Klasifikasi <i>RE-WLR</i> pada Data <i>Training</i> dan Data <i>Testing</i>	85
Tabel 4.4 Data Jumlah Kelas Positif dan Negatif pada Data <i>Training</i> dan Data <i>Testing</i> Stratifikasi.....	86
Tabel 4.5 Rata-rata Kinerja Klasifikasi <i>RE-WLR</i> (dengan Stratifikasi) untuk Klasifikasi Desa Tertinggal di Jawa Timur.....	87
Tabel 4.6 Perbandingan Kinerja Klasifikasi <i>RE-WLR</i> pada Data <i>Training</i> dan Data <i>Testing</i> Stratifikasi.....	88
Tabel 4.7 Perbandingan Kinerja Klasifikasi <i>TR-IRLS</i> dan <i>RE-WLR</i> pada λ Optimal.....	89
Tabel 4.8 Perbandingan Kinerja Klasifikasi <i>TR-IRLS</i> dan <i>RE-WLR</i> pada λ Optimal (Data Stratifikasi).....	91

DAFTAR LAMPIRAN

Lampiran 1 Data Indeks Pembangunan Desa (IPD) Jawa Timur 2014.....	101
Lampiran 2. Data yang digunakan untuk membentuk model klasifikasi.....	102
Lampiran 3. Klasifikasi Desa Tertinggal dan Tidak Tertinggal menurut Kabupaten/Kota di Jawa Timur Tahun 2014.....	103
Lampiran 4. Jumlah dan Persentase Desa menurut Kategori Variabel Penelitian.....	105
Lampiran 5. Syntax R untuk klasifikasi dengan <i>TR-IRLS</i> dan <i>RE-WLR</i>	108
Lampiran 6. Contoh output dari klasifikasi dengan <i>TR-IRLS</i> dan <i>RE-WLR</i> menggunakan R.....	118
Lampiran 7. <i>Confusion Matrix</i> Model <i>TR-IRLS</i> dan <i>RE-WLR</i> Data <i>Training</i>	127
Lampiran 8. <i>Confusion Matrix</i> Model <i>TR-IRLS</i> dan <i>RE-WLR</i> Data <i>Testing</i>	130
Lampiran 9. <i>Accuracy</i> , <i>Sensitivity</i> , <i>Specificity</i> , AUC dan <i>G-mean</i> Hasil Klasifikasi dengan <i>TR-IRLS</i> dan <i>RE-WLR</i> Data <i>Training</i>	133
Lampiran 10. <i>Accuracy</i> , <i>Sensitivity</i> , <i>Specificity</i> , AUC dan <i>G-mean</i> Hasil Klasifikasi dengan <i>TR-IRLS</i> dan <i>RE-WLR</i> Data <i>Testing</i>	135
Lampiran 11. <i>Confusion Matrix</i> model <i>TR-IRLS</i> dan <i>RE-WLR</i> Data <i>Training</i> Stratifikasi.....	137
Lampiran 12. <i>Confusion Matrix</i> Model <i>TR-IRLS</i> dan <i>RE-WLR</i> Data <i>Testing</i> Stratifikasi.....	140
Lampiran 13. <i>Accuracy</i> , <i>Sensitivity</i> , <i>Specificity</i> , AUC dan <i>G-mean</i> Hasil Klasifikasi dengan <i>TR-IRLS</i> dan <i>RE-WLR</i> Data <i>Training</i> Stratifikasi.....	143
Lampiran 14. <i>Accuracy</i> , <i>Sensitivity</i> , <i>Specificity</i> , AUC dan <i>G-mean</i> Hasil Klasifikasi dengan <i>TR-IRLS</i> dan <i>RE-WLR</i> Data <i>Testing</i> Stratifikasi.....	145
Lampiran 15. Perbandingan Box Plot Ukuran Ketepatan Klasifikasi <i>TR-IRLS</i> dan <i>RE-WLR</i> pada λ Optimal (Partisi Data Random)	147
Lampiran 16. Uji Wilcoxon dengan SPSS (Partisi Data Random).....	148

Lampiran 17. Perbandingan Box Plot Ukuran Ketepatan Klasifikasi TR - $IRLS$ dan $RE-WLR$ pada λ optimal (Partisi Data Stratifikasi)	149
Lampiran 18. Uji Wilcoxon dengan SPSS (Partisi Data Stratifikasi).....	150
Lampiran 19. Urutan Signifikansi Variabel Berdasarkan Nilai $\tilde{\beta}_j$	151

BAB 1

PENDAHULUAN

1.1 Latar Belakang

Klasifikasi data merupakan proses penting dalam bidang *data mining*. Klasifikasi adalah proses menemukan sekumpulan model atau fungsi yang menggambarkan dan membedakan konsep atau kelas-kelas data, dengan tujuan agar model tersebut dapat digunakan untuk memprediksi kelas dari suatu objek atau data (Han dan Kamber 2001).

Berbagai metode digunakan untuk klasifikasi, di antaranya Analisis Diskriminan, Regresi Logistik, dan beberapa metode klasifikasi lain dengan pendekatan *computational programming*, misalnya, *Artificial Neural Network (ANN)*, *Naive Bayes*, *Classification Adaptive Regression Tree (CART)* dan *Support Vector Machine (SVM)* (Witten, Frank, dan Hall, 2011). Selain Analisis Diskriminan, Regresi Logistik adalah metode klasifikasi yang sering digunakan.

Keuntungan menggunakan Regresi Logistik adalah metode tersebut telah dipelajari secara ekstensif. Regresi Logistik juga digunakan untuk mengetahui pengaruh beberapa peubah penjelas yang bersifat numerik atau kategorik terhadap peubah terikat yang bersifat kategorik (Kutner, Nachtsheim dan Neter, 2004). Regresi Logistik telah terbukti menjadi *classifier* yang tangguh dengan memberikan probabilitas dan mencakup masalah klasifikasi multi kelas (Hastie, Tibshirani, dan Friedman, 2001; Karsmaker, Pelckmans, dan Suykens, 2007). Informasi probabilitas dari Regresi Logistik memberikan kemudahan interpretasi dalam model klasifikasi. Selanjutnya, Regresi Logistik hanya memerlukan pemecahan masalah *unconstrained optimization*. Oleh karena itu, dengan algoritma yang tepat, waktu perhitungan bisa lebih cepat dibandingkan metode lain, seperti *Support Vector Machine (SVM)*, yang memerlukan pemecahan *constrained quadratic optimization* (Maalouf dan Siddiqi, 2014).

Regresi Logistik adalah metode klasifikasi linier yang merupakan *machine learning* dan alat *data mining* yang sangat penting. Dibandingkan dengan teknik klasifikasi lainnya, seperti metode kernel yang mentransformasikan data kedalam

ruang dimensi yang lebih tinggi, *linear classifiers* diimplementasikan langsung pada data dalam ruang dimensi asal. Keuntungan utama dari *linear classifiers* adalah prosedur *training* dan *testing* yang efisien, terutama ketika diimplementasikan pada data besar dan berdimensi tinggi (Yuan, Ho, dan Lin, 2012).

Salah satu permasalahan dalam klasifikasi data adalah komposisi data yang tidak seimbang (*imbalanced data*). Pada klasifikasi biner atau dua kelas, salah satu kelas memiliki jumlah sampel lebih besar dari kelas lainnya. Kelas data yang banyak merupakan kelas mayoritas sedangkan kelas data yang sedikit atau jarang terjadi merupakan kelas minoritas. Permasalahan yang terjadi adalah akurasi prediksi yang baik terhadap kelas data *training* yang banyak (kelas mayoritas) dan akurasi prediksi yang buruk untuk kelas data *training* yang sedikit (kelas minoritas). Atau dengan kata lain, *classifier* cenderung memprediksi kelas yang memiliki komposisi data lebih besar. Secara umum, *classifier* menghasilkan ketepatan klasifikasi yang kurang baik.

Permasalahan *imbalanced data* terjadi pada berbagai bidang antara lain klasifikasi teks (Lewis dan Catlett, 1994), deteksi tumpahan minyak dari pencitraan satelit (Kubat, Matwin dan Holte, 1998), diagnosa medis (Kononenko, 2001), deteksi penipuan kartu kredit (Wu dan Chang, 2003), telekomunikasi (Hilas, 2009), data medis kanker (Sain dan Purnami, 2013), prediksi kebangkrutan perusahaan (Haerdle, Prastyo, dan Hafner, 2014; Haerdle dan Prastyo, 2014), dan lain-lain.

Terdapat berbagai metode untuk klasifikasi *imbalanced data* yang terbagi menjadi tiga kategori utama yaitu pendekatan algoritma, pendekatan *data preprocessing* dan pendekatan seleksi fitur. Masing-masing teknik ini memiliki kelebihan dan kekurangan (Longadge, Dongre, dan Malik, 2013). Pada pendekatan algoritma, sebagian besar algoritma ditujukan untuk kelas yang seimbang, sehingga jika diterapkan pada kelas yang tidak seimbang maka ketepatan klasifikasi secara umum tidak baik.

Beberapa pengembangan metode Regresi Logistik telah dilakukan untuk meningkatkan ketepatan klasifikasi pada *imbalanced data*, di antaranya dengan pendekatan kernel dan non kernel (linier). Dengan pendekatan kernel, Maalouf

dan Trafalis (2010) mengembangkan metode *Rare Event Weighted Kernel Logistic Regression (RE-WKLR)* yang sesuai untuk data berukuran kecil sampai sedang. Pendekatan kernel juga digunakan Rahayu (2012) dalam mengembangkan metode *AdaBoost Newton Truncated Regularized Weighted Kernel Logistic Regression (AB-WKLR)* dan *AdaBoost NTR Weighted Regularized Logistic Regression (AB-WLR)* yang secara signifikan berhasil meningkatkan performansi akurasi dan stabilitas *general classifiers* pada *NTR-KLR* dan *NTR-LR*. Selanjutnya, pendekatan non kernel (linier) digunakan Maalouf dan Siddiqi (2014) dalam mengembangkan metode *Rare Event Weighted Logistic Regression (RE-WLR)* untuk klasifikasi *imbalanced data* pada data berskala besar. Penelitian tersebut menghasilkan kesimpulan bahwa *RE-WLR* memiliki performansi yang lebih baik dibandingkan *Truncated-Regularized Iteratively Re-weighted Least Squares (TR-IRLS)*. *TR-IRLS* merupakan metode yang dikembangkan oleh Komarek dan Moore (2005) untuk Regresi Logistik pada data berskala besar.

Berdasarkan keutamaan *linear classifier*, maka penelitian ini menerapkan metode *Rare Event Weighted Logistic Regression (RE-WLR)* untuk klasifikasi *imbalanced data* dengan studi kasus klasifikasi desa tertinggal di Provinsi Jawa Timur. Jawa Timur merupakan provinsi dengan jumlah desa terbesar kedua di Indonesia. Besarnya jumlah desa tersebut memerlukan perencanaan dan pengambilan kebijakan yang tepat dalam rangka pembangunan desa. Berdasarkan data yang dipublikasikan oleh Badan Perencanaan Pembangunan Nasional dan Badan Pusat Statistik (2015), 208 dari 7.721 desa di 38 kabupaten/kota di Jawa Timur atau sebanyak 2,67% desa masuk kategori desa tertinggal.

Desa Tertinggal adalah desa yang belum terpenuhi SPM desa pada aspek kebutuhan sosial dasar, infrastruktur dasar, sarana dasar, pelayanan umum, dan penyelenggaraan pemerintahan (Bappenas dan BPS, 2015). Penetapan status desa tertinggal diharapkan dapat menjadi manifestasi suatu ukuran yang mengidentifikasi daerah kantong-kantong kemiskinan (BPS, 2005).

Salah satu aspek terpenting untuk mendukung strategi penanggulangan kemiskinan adalah tersedianya data kemiskinan yang akurat dan tepat sasaran. Data kemiskinan yang baik dapat digunakan untuk mengevaluasi kebijakan pemerintah terhadap kemiskinan, membandingkan kemiskinan antar waktu dan

daerah, serta menentukan target penduduk miskin dengan tujuan untuk memperbaiki kondisi suatu daerah (BPS, 2008). Upaya lainnya untuk mengentaskan kemiskinan juga diwujudkan dengan Inpres No.5 Tahun 1993, tentang peningkatan penanggulangan kemiskinan, sebagai program khusus mempercepat upaya mengatasi masalah kemiskinan yang diarahkan secara langsung pada penduduk miskin di desa tertinggal (Bappenas, 1993).

Pada tahun 1993, 1994, dan 1995 BPS telah melakukan pengklasifikasian desa tertinggal sebagai proksi identifikasi daerah kantong-kantong kemiskinan. Data yang digunakan untuk penentuan desa tertinggal pada tahun 1993 adalah data PODES dengan menggunakan 25 variabel untuk daerah perkotaan dan 27 variabel untuk daerah pedesaan. Identifikasi status ketertinggalan desa pada tahun 1994 menggunakan 17 variabel untuk daerah perkotaan dan 18 variabel untuk daerah pedesaan. Klasifikasi desa tertinggal dilakukan kembali pada tahun 2003 dengan menggunakan data PODES ST03 dan SUSENAS 2002, dengan menggunakan 45 variabel.

Variabel yang digunakan dalam penetapan klasifikasi desa tertinggal tidak dapat menggunakan variabel yang sama untuk setiap tahunnya. Hal ini disebabkan karena adanya perubahan status desa, pemekaran atau penggabungan desa, pengembangan wilayah dengan membuka kawasan baru, perubahan (penambahan/pengurangan) variabel yang dikumpulkan dalam PODES, dan perubahan variasi kesejahteraan keluarga di desa tersebut. Dengan demikian indikator-indikator desa tertinggal masih perlu dikembangkan dan dievaluasi dalam penelitian-penelitian selanjutnya (Mulya, 2005).

Beberapa penelitian mengenai identifikasi desa tertinggal telah dilakukan dengan menggunakan metode seleksi variabel yang diduga menjadi faktor penentu status ketertinggalan desa. Salah satu metode yang digunakan antara lain uji validitas variabel yang mempunyai hubungan dengan rata-rata pengeluaran per kapita penduduk dan metode *stepwise discriminant analysis* (BPS, 2005).

Penelitian-penelitian lain mengenai desa tertinggal juga telah dilakukan. Hidayat (2008), melakukan pemodelan desa tertinggal di Jawa Barat Tahun 2005 dengan pendekatan *Multivariate Adaptive Regression Splines (MARS)*. Penelitian lain mengenai ketertinggalan daerah adalah Evaluasi Ketertinggalan Daerah

Dengan Analisis Diskriminan (Djuraidah, 2009) dan Penggunaan *Geographically Weighted Regression-Kriging* untuk Klasifikasi Desa Tertinggal (Dimulyo, 2009). Wahyuningsih (2009) mengidentifikasi faktor-faktor penentu desa tertinggal dan mengkaji keterkaitan antar faktor penentu desa tertinggal di provinsi Lampung dengan menggunakan metode *Structural Equation Modeling (SEM)*. Hanifah (2010), melakukan pengelompokan desa tertinggal dengan metode *Latent Class Cluster* untuk variabel indikator bertipe campuran. Nooraeni (2015) melakukan pengelompokan desa berdasarkan indikator ketertinggalan desa dengan Metode Kombinasi Algoritma *Cluster K-Prototype* dan Algoritma Genetika.

1.2 Rumusan Masalah

Berdasarkan uraian latar belakang di atas, maka masalah dalam penelitian ini dapat dirumuskan sebagai berikut:

1. Bagaimana bentuk penaksir parameter pada model *Rare Event Weighted Logistic Regression (RE-WLR)*?
2. Bagaimana penerapan *RE-WLR* untuk klasifikasi desa tertinggal di Provinsi Jawa Timur?
3. Bagaimana hasil perbandingan ketepatan metode klasifikasi antara *RE-WLR* dan *TR-IRLS* dalam pengklasifikasian *imbalanced data* dengan studi kasus klasifikasi desa tertinggal di Provinsi Jawa Timur?

1.3 Tujuan Penelitian

Berdasarkan rumusan masalah yang telah diuraikan diatas, maka tujuan yang ingin dicapai dalam penelitian ini adalah:

1. Memperoleh bentuk penaksir parameter pada model *RE-WLR*.
2. Menerapkan *RE-WLR* untuk klasifikasi desa tertinggal di Provinsi Jawa Timur.
3. Membandingkan tingkat ketepatan klasifikasi antara *RE-WLR* dan *TR-IRLS* dalam pengklasifikasian *imbalanced data* dengan studi kasus klasifikasi desa tertinggal di Provinsi Jawa Timur.

1.4 Manfaat Penelitian

Manfaat yang ingin dicapai dalam penelitian ini adalah sebagai berikut:

1. Menambah wawasan keilmuan mengenai permasalahan dan penanganan pada *imbalanced data* dengan penerapan *RE-WLR*.
2. Memberikan wawasan tentang alternatif metode untuk klasifikasi desa tertinggal.

1.5 Batasan Penelitian

Pada penelitian ini, masalah dibatasi hanya pada klasifikasi biner dengan kasus klasifikasi desa tertinggal di Provinsi Jawa Timur Tahun 2014.

BAB 2

TINJAUAN PUSTAKA

2.1 Regresi Logistik / *Logistic Regression (LR)*

LR adalah metode yang dapat digunakan untuk menyatakan hubungan variabel respon yang bersifat kategorik (dalam bentuk dikotomis maupun polycotomis) dengan variabel prediktor baik yang bersifat kategorik maupun kontinu (Agresti, 1996). Tujuan regresi logistik pada dasarnya adalah pengklasifikasian setiap kejadian dalam suatu pengamatan ke dalam kelas positif dan negatif.

Misalkan $\mathbf{x}_i \in R^{k+1}$ adalah vektor untuk setiap kasus di \mathbf{X} dimana $i = 1 \dots n$, $\boldsymbol{\beta}$ adalah vektor parameter dan \mathbf{y} adalah vektor respon biner yang bernilai 0 (kelas mayoritas/negatif) atau 1 (kelas minoritas/positif). Pada dasarnya, regresi logistik dibangun untuk variabel prediktor kontinyu ($x \in R$). Pada perkembangannya, regresi logistik juga diterapkan pada variabel prediktor kategori.

$$\mathbf{y} = \begin{bmatrix} y_1 \\ y_2 \\ \vdots \\ y_i \\ \vdots \\ y_n \end{bmatrix}, \quad \mathbf{X} = \begin{bmatrix} 1 & x_{11} & x_{12} & \cdots & x_{1k} \\ 1 & x_{21} & x_{22} & \cdots & x_{2k} \\ \vdots & \vdots & \vdots & & \vdots \\ 1 & x_{i1} & x_{i2} & \cdots & x_{ik} \\ \vdots & \vdots & \vdots & & \vdots \\ 1 & x_{n1} & x_{n2} & \cdots & x_{nk} \end{bmatrix} = \begin{bmatrix} \mathbf{x}_1^T \\ \mathbf{x}_2^T \\ \vdots \\ \mathbf{x}_i^T \\ \vdots \\ \mathbf{x}_n^T \end{bmatrix}, \quad \boldsymbol{\beta} = \begin{bmatrix} \beta_0 \\ \beta_1 \\ \beta_2 \\ \vdots \\ \beta_k \end{bmatrix}.$$

Variabel respon (y) mengikuti distribusi Bernoulli dengan peluang sukses sebesar p yang mempunyai fungsi probabilitas:

$$f(y_i) = (p_i)^{y_i}(1 - p_i)^{(1-y_i)}, \quad y_i = \{0, 1\} \quad (2.1)$$

Menurut Hosmer dan Lemeshow (2000), fungsi logistik yang digunakan untuk memodelkan \mathbf{x}_i dengan nilai ekspektasi y_i nya yaitu

$$E(y_i | \mathbf{x}_i, \boldsymbol{\beta}) = p_i = \frac{e^{\mathbf{x}_i^T \boldsymbol{\beta}}}{1 + e^{\mathbf{x}_i^T \boldsymbol{\beta}}} = \frac{1}{1 + e^{-\mathbf{x}_i^T \boldsymbol{\beta}}}. \quad (2.2)$$

LR menggunakan fungsi transformasi logit yang merupakan nilai logaritma *odds* dari respon positif yaitu

$$\eta_i = \ln \left(\frac{p_i}{1 - p_i} \right) = \mathbf{x}_i^T \boldsymbol{\beta}, \quad (2.3)$$

Dalam bentuk matriks, fungsi transformasi logit adalah

$$\boldsymbol{\eta} = \mathbf{X}\boldsymbol{\beta}, \quad (2.4)$$

dimana

$\boldsymbol{\eta}$ = vektor logit dengan ukuran $n \times 1$,

\mathbf{X} = matriks dari variabel prediktor dengan ukuran $n \times (k+1)$,

$\boldsymbol{\beta}$ = vektor parameter dengan ukuran $(k+1) \times 1$.

1.2 Regularized Logistic Regression

Dengan asumsi bahwa pengamatan independen, fungsi *Likelihood* dari LR adalah

$$L(\boldsymbol{\beta}) = \prod_{i=1}^n (p_i)^{y_i} (1 - p_i)^{(1-y_i)} = \prod_{i=1}^n \left(\frac{e^{\mathbf{x}_i^T \boldsymbol{\beta}}}{1 + e^{\mathbf{x}_i^T \boldsymbol{\beta}}} \right)^{y_i} \left(\frac{1}{1 + e^{\mathbf{x}_i^T \boldsymbol{\beta}}} \right)^{(1-y_i)}. \quad (2.5)$$

Fungsi *regularized log likelihood* didefinisikan sebagai berikut

$$\ln L(\boldsymbol{\beta}) = \sum_{i=1}^n \left(y_i \ln \left(\frac{e^{\mathbf{x}_i^T \boldsymbol{\beta}}}{1 + e^{\mathbf{x}_i^T \boldsymbol{\beta}}} \right) + (1 - y_i) \ln \left(\frac{1}{1 + e^{\mathbf{x}_i^T \boldsymbol{\beta}}} \right) \right) - \frac{\lambda}{2} \|\boldsymbol{\beta}\|^2 \quad (2.6)$$

$$= \sum_{i=1}^n \ln \left(\frac{e^{-y_i \mathbf{x}_i^T \boldsymbol{\beta}}}{1 + e^{\mathbf{x}_i^T \boldsymbol{\beta}}} \right) - \frac{\lambda}{2} \|\boldsymbol{\beta}\|^2, \quad (2.7)$$

dimana $\|\boldsymbol{\beta}\| = \sqrt{\beta_0^2 + \beta_1^2 + \beta_2^2 + \dots + \beta_k^2}$. Regularisasi (penalti) $\frac{\lambda}{2} \|\boldsymbol{\beta}\|^2$ ditambahkan untuk memperoleh generalisasi yang lebih baik dalam mengatasi *over fitting*.

Untuk output biner, fungsi *loss* atau penyimpangan DEV adalah *negative log likelihood* dengan rumus sebagai berikut (Komarek, 2004; Hosmer dan Lemeshow, 2000)

$$DEV(\hat{\boldsymbol{\beta}}) = -2 \ln L(\hat{\boldsymbol{\beta}}). \quad (2.8)$$

2.3 *Truncated-Regularized Iteratively Re-weighted Least Squares (TR-IRLS)*

TR-IRLS diperkenalkan oleh Komarek dan Moore (2005) yang mengkombinasikan regularisasi, metode *Truncated-Newton*, dan *Iteratively Re-weighted Least Square (IRLS)*. Pada penelitian yang dilakukan oleh Komarek dan Moore tersebut, algoritma *TR-IRLS* digunakan pada *LR* dan menghasilkan performansi yang lebih baik daripada linear *SVM* dan *RBF SVM*.

Metode yang digunakan untuk mengestimasi parameter regresi logistik adalah MLE (*Maximum Likelihood Estimator*). Estimasi maksimum likelihood merupakan pendekatan dari estimasi WLS (*Weighted Least Square*), dimana matriks pembobotnya berubah setiap iterasi. Proses perhitungan estimasi ini disebut juga sebagai *Iteratively Reweighted Least Square (IRLS)*.

IRLS menggunakan metode iterasi Newton-Raphson untuk memecahkan persamaan nilai LR. Gradien dan Hessian diperoleh dengan menurunkan *Regularized Log-Likelihood* terhadap β .

$$\frac{\partial}{\partial \beta} \ln L(\beta) = \mathbf{X}^T(\mathbf{y} - \mathbf{p}) - \lambda \beta = 0, \quad (2.9)$$

$$\frac{\partial^2}{\partial^2 \beta} \ln L(\beta) = -\mathbf{X}^T \mathbf{V} \mathbf{X} - \lambda \mathbf{I} = 0, \quad (2.10)$$

dimana

$$v_i = p_i(1 - p_i), \quad (2.11)$$

dan

$$\mathbf{V} = \text{diag}(v_1, \dots, v_n). \quad (2.12)$$

Subproblem *WLS* adalah

$$(\mathbf{X}^T \mathbf{V} \mathbf{X} + \lambda \mathbf{I}) \hat{\beta}^{(c+1)} = \mathbf{X}^T \mathbf{V} \mathbf{z}^{(c)}. \quad (2.13)$$

Update Newton-Raphson untuk $\hat{\beta}$ pada iterasi ke $(c+1)$ yaitu

$$\hat{\beta}^{(c+1)} = (\mathbf{X}^T \mathbf{V} \mathbf{X} + \lambda \mathbf{I})^{-1} \mathbf{X}^T \mathbf{V} \mathbf{z}^{(c)}, \quad (2.14)$$

dimana

$$\mathbf{z}^{(c)} = \mathbf{X} \hat{\beta}^{(c)} + \mathbf{V}^{-1}(\mathbf{y} - \mathbf{p}) \quad (2.15)$$

adalah respon yang disesuaikan.

RLR mempunyai problem numerik yaitu dalam mendapatkan invers dari matriks Hessian berukuran besar yang mengakibatkan pemrosesan menjadi lama karena ukuran yang besar. Salah satu cara yang digunakan untuk mengatasi

problem numerik tersebut adalah penggunaan metode *Truncated Newton* dengan Linier CG sebagai algoritma *truncated inner*.

Pada dasarnya metode *Truncated Newton* terdiri dari metode iterasi ganda yaitu iterasi luar dari problem optimasi non linier (seperti *MLE*) dan iterasi dalam untuk persamaan Newton. Sebelum persamaan Newton didapatkan, iterasi akan dihentikan atau *truncated*/terputus. Pada setiap iterasi estimasi solusi diperbarui dengan pendekatan persamaan Newton yang menggunakan algoritma iterasi. Untuk data dalam skala besar, metode *Truncated Newton* terbukti powerful, fleksibel dan lebih mudah diterapkan (Rahayu, dkk., 2012).

Algoritma 1. LR MLE menggunakan IRLS (Maalouf, 2009)

Input : $\mathbf{X}, \mathbf{y}, \hat{\boldsymbol{\beta}}^{(0)}$

Output : $\hat{\boldsymbol{\beta}}$

Begin

$c = 0$

Do While $\left| \frac{DEV^{(c)} - DEV^{(c+1)}}{DEV^{(c+1)}} \right| > \varepsilon_1$ dan $c \leq \text{Max IRLS Iterations}$

For $i = 1$ to n

Hitung probabilitas $\hat{p}_i = \frac{e^{\mathbf{x}_i^T \hat{\boldsymbol{\beta}}}}{1 + e^{\mathbf{x}_i^T \hat{\boldsymbol{\beta}}}}$

Hitung pembobot $v_i = \hat{p}_i(1 - \hat{p}_i)$

Hitung *adjusted response* $z_i = \mathbf{x}_i \hat{\boldsymbol{\beta}}^{(c)} + \frac{(y_i - \hat{p}_i)}{\hat{p}_i(1 - \hat{p}_i)}$

End For

$\mathbf{V} = \text{diag}(v_1, \dots, v_n)$

Hitung $\hat{\boldsymbol{\beta}}$ via algoritma (2) $(\mathbf{X}^T \mathbf{V} \mathbf{X} + \lambda \mathbf{I}) \hat{\boldsymbol{\beta}}^{(c+1)} = \mathbf{X}^T \mathbf{V} \mathbf{z}^{(c)}$

$c = c + 1$

End while

End

Dimana *max IRLS iterations* = 30 dan $\varepsilon_1 = 0.01$

Algoritma 2. Linear CG untuk menghitung $\hat{\beta}$, $A = X^T V X + \lambda I$, $b = X^T V z$

(Rahayu, dkk., 2012)

Input : $A, b, \hat{\beta}^{(0)}$

Output : $\hat{\beta}$ sehingga $A\hat{\beta} = b$

Begin

 Inisialisasi $r^{(0)} = b - A\hat{\beta}^{(0)}$

$d^{(0)} = r^{(0)}$

$t = 0$

 Do While $\|r^{(t)}\|^2 > \varepsilon_2$ dan $t \leq \text{Max CG Iterations}$

 Hitung panjang langkah optimal $s^{(t)} = \frac{r^{T(t)}r^{(t)}}{d^{T(t)}Ad^{(t)}}$

 Hitung perkiraan solusi $\hat{\beta}^{(t+1)} = \hat{\beta}^{(t)} + s^{(t)}d^{(t)}$

 Update residual $r^{(t+1)} = r^{(t)} - s^{(t)}Ad^{(t)}$

 Update *A-conjugacy* $a^{(t)} = \frac{r^{T(t+1)}r^{(t+1)}}{r^{T(t)}r^{(t)}}$

 Update arah pencarian $d^{(t+1)} = r^{(t+1)} + a^{(t)}d^{(t)}$

$t = t + 1$

 End while

End

Dimana *max CG iterations* = 200 dan $\varepsilon_2 = 0.005$

2.4 Rare Event Weighted Logistic Regression (RE-WLR)

Maalouf dan Siddiqi (2014) memperkenalkan *Rare Event Weighted Logistic Regression (RE-WLR)* sebagai penerapan dari *LR rare event correction* pada algoritma *TR-IRLS*. Algoritma *RE-WLR* berdasarkan algoritma *Rare Event Weighted Kernel Logistic Regression (RE-WKLR)* yang dikembangkan Maalouf dan Trafalis (2011).

Rare event adalah kejadian yang frekuensi terjadinya lebih kecil daripada frekuensi kejadian pada umumnya. Pada kasus *imbalanced data* atau *rare event*, King dan Zeng (2001) mengusulkan *weighting* sebagai prosedur alternatif untuk mengkoreksi estimasi pemilihan sampel y .

Selanjutnya fungsi *log likelihood* untuk *LR* dapat ditulis sebagai

$$\ln L(\boldsymbol{\beta}|\mathbf{y}, \mathbf{X}) = \sum_{i=1}^n \frac{Q_i}{H_i} \ln P(y_i|\mathbf{x}_i, \boldsymbol{\beta}) \quad (2.16)$$

$$= \sum_{i=1}^n \frac{Q_i}{H_i} \ln \left(\frac{e^{-y_i \mathbf{x}_i^T \boldsymbol{\beta}}}{1 + e^{\mathbf{x}_i^T \boldsymbol{\beta}}} \right) \quad (2.17)$$

$$= \sum_{i=1}^n w_i \ln \left(\frac{e^{-y_i \mathbf{x}_i^T \boldsymbol{\beta}}}{1 + e^{\mathbf{x}_i^T \boldsymbol{\beta}}} \right), \quad (2.18)$$

$$\text{dimana } w_i = \frac{Q_i}{H_i} = \left(\frac{\tau}{\bar{y}} \right) y_i + \left(\frac{1-\tau}{1-\bar{y}} \right) (1 - y_i), \quad (2.19)$$

dengan Q adalah proporsi kejadian dalam sampel, H proporsi kejadian dalam populasi, \bar{y} adalah proporsi kejadian sukses (*event*) dalam sampel dan τ adalah penduga proporsi kejadian sukses dalam populasi.

Menurut McCullagh dan Nelder (1989), dan kemudian Cordeiro dan McCullagh (1991), vektor bias diberikan oleh persamaan berikut

$$\text{bias}(\hat{\boldsymbol{\beta}}) = (\mathbf{X}^T \mathbf{V} \mathbf{X})^{-1} \mathbf{X}^T \mathbf{V} \boldsymbol{\xi}, \quad (2.20)$$

$$\text{dimana } \xi_i = Q_{ii} \left(\hat{p}_i - \frac{1}{2} \right),$$

dan Q_{ii} adalah elemen diagonal dari $\mathbf{Q} = \mathbf{X}(\mathbf{X}^T \mathbf{V} \mathbf{X})^{-1} \mathbf{X}^T$.

Estimator dengan bias terkoreksi menjadi

$$\tilde{\boldsymbol{\beta}} = \hat{\boldsymbol{\beta}} - \mathbf{B}(\hat{\boldsymbol{\beta}}). \quad (2.21)$$

Perumusan *Weighted Logistic Regression* dari King dan Zeng (2001) diterapkan pada (2.5), sehingga fungsi *Weighted Likelihood* dapat ditulis sebagai

$$L_w(\boldsymbol{\beta}) = \prod_{i=1}^n (p_i)^{w_1 y_i} (1 - p_i)^{w_0 (1 - y_i)}, \quad (2.22)$$

dimana $w_1 = \frac{\tau}{\bar{y}}$ dan $w_0 = \frac{1-\tau}{1-\bar{y}}$, dengan w_1 adalah pembobot untuk $y_i = 1$

dan w_0 adalah pembobot untuk $y_i = 0$.

Selanjutnya probabilitas kejadian sukses dengan pembobot adalah

$$p_i^* = \left(\frac{1}{1 + e^{-\eta_i}} \right)^{w_1} = p_i^{w_1}, \quad (2.23)$$

dimana $\eta_i = \mathbf{x}_i^T \boldsymbol{\beta}$.

$$p_i^{*'} = w_1 p_i^{w_1} (1 - p_i), \quad (2.24)$$

$$p_i^{*''} = w_1 p_i^{w_1} (1 - p_i) (w_1 - (1 + w_1) p_i). \quad (2.25)$$

Vektor Bias untuk WLR adalah

$$\mathbf{B}(\hat{\boldsymbol{\beta}}) = (\mathbf{X}^T \mathbf{D} \mathbf{X})^{-1} \mathbf{X}^T \mathbf{D} \boldsymbol{\xi}, \quad (2.26)$$

dimana elemen ke- i dari vector $\boldsymbol{\xi}$ adalah

$$\xi_i = -0,5 \left(\frac{p_i^{*''}}{p_i^{*'}} \right) Q_{ii} = 0,5 Q_{ii} ((1 + w_1) p_i - w_1), \quad (2.27)$$

dengan Q_{ii} adalah elemen diagonal dari matriks kovarians \mathbf{Q} dimana sekarang menjadi

$$\mathbf{Q} = \mathbf{X}(\mathbf{X}^T \mathbf{D} \mathbf{X})^{-1} \mathbf{X}^T, \quad (2.28)$$

dan

$$\mathbf{D} = \text{diag} (v_i w_i) \text{ untuk } i = 1 \dots n. \quad (2.29)$$

Estimator dengan bias terkoreksi menjadi

$$\tilde{\boldsymbol{\beta}} = \hat{\boldsymbol{\beta}} - \mathbf{B}(\hat{\boldsymbol{\beta}}). \quad (2.30)$$

Untuk *WLR*, gradien dan Hessian diperoleh dengan menurunkan *Regularized Weighted Log-Likelihood*

$$\ln L_w(\boldsymbol{\beta}) = \sum_{i=1}^n w_i \ln \frac{e^{y_i \mathbf{x}_i^T \boldsymbol{\beta}}}{1 + e^{\mathbf{x}_i^T \boldsymbol{\beta}}} - \frac{\lambda}{2} \|\boldsymbol{\beta}\|^2. \quad (2.31)$$

Dalam bentuk matriks, gradient yaitu

$$\nabla_{\boldsymbol{\beta}} \ln L_w(\boldsymbol{\beta}) = \mathbf{X}^T \mathbf{W}(\mathbf{y} - \mathbf{p}) - \lambda \boldsymbol{\beta}, \quad (2.32)$$

dimana $\mathbf{W} = \text{diag} (w_i)$ dan \mathbf{p} adalah vektor probabilita. Hessian dari $\boldsymbol{\beta}$ yaitu

$$\nabla_{\boldsymbol{\beta}}^2 \ln L_w(\boldsymbol{\beta}) = -\mathbf{X}^T \mathbf{D} \mathbf{X} - \lambda \mathbf{I}. \quad (2.33)$$

Update Newton-Raphson untuk $\boldsymbol{\beta}$ pada iterasi ke $(c+1)$ yaitu

$$\hat{\boldsymbol{\beta}}^{(c+1)} = (\mathbf{X}^T \mathbf{D} \mathbf{X} + \lambda \mathbf{I})^{-1} \mathbf{X}^T \mathbf{D} \mathbf{z}^{(c)}, \quad (2.34)$$

$$\text{dimana } \mathbf{z}^{(c)} = \mathbf{X} \hat{\boldsymbol{\beta}}^{(c)} + \mathbf{V}^{-1}(\mathbf{y} - \mathbf{p}) \quad (2.35)$$

adalah variabel respon yang disesuaikan.

Subproblem *WLS* adalah

$$(\mathbf{X}^T \mathbf{D} \mathbf{X} + \lambda \mathbf{I}) \hat{\boldsymbol{\beta}}^{(c+1)} = \mathbf{X}^T \mathbf{D} \mathbf{z}^{(c)}. \quad (2.36)$$

Untuk memecahkan sub problem *WLS*, maka digunakan metode *Truncated Newton* dengan Linier CG sebagai algoritma *truncated inner*. Selanjutnya, prediksi peluang dilakukan dengan memasukkan penaksir terkoreksi ($\tilde{\boldsymbol{\beta}}$) ke dalam persamaan logit sebagai berikut:

$$\tilde{p}_i = \frac{e^{\mathbf{x}_i^T \tilde{\boldsymbol{\beta}}}}{1 + e^{\mathbf{x}_i^T \tilde{\boldsymbol{\beta}}}}. \quad (2.37)$$

Algoritma 3. WLR MLE menggunakan IRLS (Maalouf dan Siddiqi, 2014)

Input : $\mathbf{X}, \mathbf{y}, \hat{\boldsymbol{\beta}}^{(0)}, w_1, w_0$

Output : $\hat{\boldsymbol{\beta}}, \mathbf{B}(\hat{\boldsymbol{\beta}}), \tilde{\boldsymbol{\beta}}, \tilde{p}_i$

Begin

$c = 0$

Do While $\left| \frac{DEV^{(c)} - DEV^{(c+1)}}{DEV^{(c+1)}} \right| > \varepsilon_1$ dan $c \leq \text{Max IRLS Iterations}$

For $i=1$ to n

Hitung probabilitas $\hat{p}_i = \frac{e^{\mathbf{x}_i^T \hat{\boldsymbol{\beta}}}}{1 + e^{\mathbf{x}_i^T \hat{\boldsymbol{\beta}}}}$

Hitung varians $v_i = \hat{p}_i(1 - \hat{p}_i)$

Hitung pembobot $w_i = w_1 y_i + w_0 (1 - y_i)$

Hitung *adjusted response* $z_i = \mathbf{x}_i \hat{\boldsymbol{\beta}}^{(c)} + \frac{(y_i - \hat{p}_i)}{\hat{p}_i(1 - \hat{p}_i)}$

End For

$\mathbf{D} = \text{diag}(v_i w_i)$

Hitung matriks kovarian $\mathbf{Q} = \mathbf{X}(\mathbf{X}^T \mathbf{D} \mathbf{X} + \lambda \mathbf{I})^{-1} \mathbf{X}^T$

$Q_{ii} = \text{diag}(\mathbf{Q})$

For $k=1$ to n

Hitung *bias response* $\xi_k = 0,5 Q_{ii_k} ((1 + w_1) \hat{p}_k - w_1)$

End For

Hitung $\hat{\boldsymbol{\beta}}$ via algoritma 4(CG) $(\mathbf{X}^T \mathbf{D} \mathbf{X} + \lambda \mathbf{I}) \hat{\boldsymbol{\beta}}^{(c+1)} = \mathbf{X}^T \mathbf{D} \mathbf{z}^{(c)}$

Hitung $\mathbf{B}(\hat{\boldsymbol{\beta}})$ via algoritma 5(CG) $(\mathbf{X}^T \mathbf{D} \mathbf{X} + \lambda \mathbf{I}) \mathbf{B}(\hat{\boldsymbol{\beta}})^{(c+1)} = \mathbf{X}^T \mathbf{D} \boldsymbol{\xi}^{(c)}$

End while

Hitung *unbiased* $\tilde{\boldsymbol{\beta}} = \hat{\boldsymbol{\beta}} - \mathbf{B}(\hat{\boldsymbol{\beta}})$

Hitung probabilitas optimal $\tilde{p}_i = \frac{e^{\mathbf{x}_i^T \tilde{\boldsymbol{\beta}}}}{1 + e^{\mathbf{x}_i^T \tilde{\boldsymbol{\beta}}}}$

End

Dimana *max IRLS iterations* = 30 dan $\varepsilon_1 = 0.01$

Algoritma 4. Linear CG untuk menghitung $\hat{\beta}$, $A = X^TDX + \lambda I$, $b = X^TDz$

(Rahayu, dkk., 2012)

Input : $A, b, \hat{\beta}^{(0)}$

Output : $\hat{\beta}$ sehingga $A\hat{\beta} = b$

Begin

Inisialisasi $r^{(0)} = b - A\hat{\beta}^{(0)}$

$d^{(0)} = r^{(0)}$

$t=0$

Do While $\|r^{(t)}\|^2 > \varepsilon_2$ dan $c \leq \text{Max CG Iterations}$

Hitung panjang langkah optimal $s^{(t)} = \frac{r^{T(t)}r^{(t)}}{d^{T(t)}Ad^{(t)}}$

Hitung perkiraan solusi $\hat{\beta}^{(t+1)} = \hat{\beta}^{(t)} + s^{(t)}d^{(t)}$

Update residual $r^{(t+1)} = r^{(t)} - s^{(t)}Ad^{(t)}$

Update *A-conjugacy* $a^{(t)} = \frac{r^{T(t+1)}r^{(t+1)}}{r^{T(t)}r^{(t)}}$

Update arah pencarian $d^{(t+1)} = r^{(t+1)} + a^{(t)}d^{(t)}$

$t = t + 1$

End while

End

Dimana *max CG iterations* = 200 dan $\varepsilon_2 = 0.005$

Algoritma 5. Linear CG untuk menghitung bias, $A = X^TDX + \lambda I$, $C = X^TD\xi$

(Rahayu, dkk., 2012)

Input : $A, C, B(\hat{\beta})^{(0)}$

Output : $B(\hat{\beta})$ sehingga $AB(\hat{\beta}) = C$

Begin

Inisialisasi $r^{(0)} = C - AB(\hat{\beta})^{(0)}$

$d^{(0)} = r^{(0)}$

$h = 0$

Do While $\|r^{(h)}\|^2 > \varepsilon_3$ dan $c \leq \text{Max CG Iterations}$

Hitung panjang langkah optimal $s^{(h)} = \frac{r^{T(h)}r^{(h)}}{d^{T(h)}Ad^{(h)}}$

Hitung perkiraan solusi $\mathbf{B}(\hat{\boldsymbol{\beta}})^{(h+1)} = \mathbf{B}(\hat{\boldsymbol{\beta}})^{(h)} + s^{(h)}\mathbf{d}^{(h)}$

Update residual $\mathbf{r}^{(h+1)} = \mathbf{r}^{(h)} - s^{(h)}\mathbf{A}\mathbf{d}^{(h)}$

Update *A-conjugacy* $a^{(h)} = \frac{\mathbf{r}^{T(h+1)}\mathbf{r}^{(h+1)}}{\mathbf{r}^{T(h)}\mathbf{r}^{(h)}}$

Update arah pencarian $\mathbf{d}^{(h+1)} = \mathbf{r}^{(h+1)} + a^{(h)}\mathbf{d}^{(h)}$

$h = h + 1$

End while

End

Dimana *max CG iterations* = 200 dan $\varepsilon_3 = 0.005$

2.5 Evaluasi Kinerja

Ukuran kinerja suatu sistem dapat diukur berdasarkan ketepatan atau akurasi klasifikasi. Misalkan ada problem klasifikasi dengan dua kelas yang memetakan satu set elemen sebagai label kelas positif/minoritas (1) dan label kelas negatif/ mayoritas (0). Metode klasifikasi yang digunakan memetakan data tersebut ke kelas yang diprediksikan. Misalkan kelas yang diprediksi disimbolkan dengan $\{0,1\}$. Hasil klasifikasi akan menghasilkan pemetaan dengan empat keluaran (Fawcett, 2006), yaitu:

- TP/True Positif (kelas sebenarnya positif dan diklasifikasikan positif)
- TN/True Negatif (kelas sebenarnya negatif dan diklasifikasikan negatif)
- FP/False Positif (kelas sebenarnya negatif dan diklasifikasikan positif)
- FN/False Negatif (kelas sebenarnya positif dan diklasifikasikan negatif)

Keempat nilai tersebut membentuk sebuah matriks yang disebut *confusion matrix*.

Tabel 2.1. *Confusion matrix* untuk perhitungan akurasi atau ketepatan klasifikasi

Kelas sebenarnya	Kelas Hasil Klasifikasi	
	Positif (1)	Negatif (0)
Positif (1)	TP	FN
Negatif (0)	FP	TN

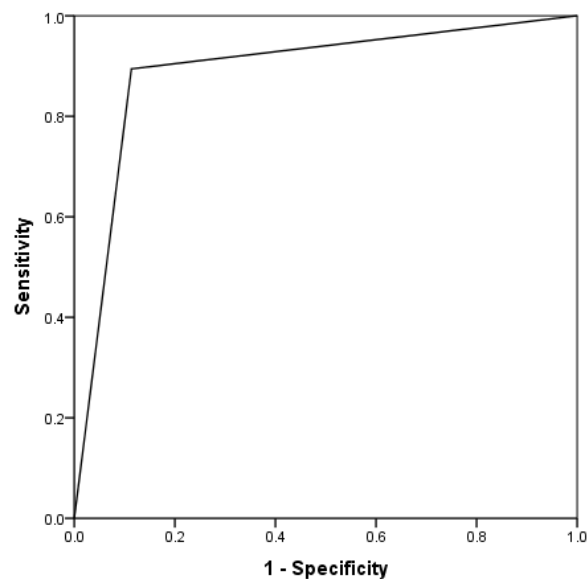
$$accuracy(\%) = \frac{TP + TN}{TP + FP + TN + FN} \times 100\% \quad (2.38)$$

Pengukuran akurasi di atas tidak mampu menunjukkan frekuensi dari FN dan FP. Hal tersebut dapat diatasi apabila dilakukan pengukuran *sensitivity* dan *specificity* dengan cara sebagai berikut

$$sensitivity(\%) = \frac{TP}{TP + FN} \times 100\% \quad (2.39)$$

$$specificity(\%) = \frac{TN}{FP + TN} \times 100\% \quad (2.40)$$

Apabila dalam evaluasi kinerja ditemukan *sensitivity* tinggi sekaligus *specificity* rendah atau sebaliknya, maka dapat digunakan kurva *Receiver Operating Characteristic (ROC)* untuk membandingkan kinerja klasifikasi. *ROC* dapat digunakan untuk menghitung nilai kesalahan dan kesuksesan suatu sistem. Kinerja tersebut dapat diukur berdasarkan perhitungan area di bawah kurva yang disebut *Area Under the ROC (AUC)*. Apabila Pengujian memiliki tingkat akurasi tinggi maka nilai *AUC* mendekati 1,00. (Fawcett, 2006)



Gambar 2.1 Kurva ROC

Selain AUC, kinerja klasifikasi dapat dihitung juga menggunakan G-mean (Kubat dan Matwin, 1997), yaitu

$$G - mean = \sqrt{sensitifity \times specificity} . \quad (2.41)$$

2.6 Uji Peringkat Bertanda Wilcoxon

Uji Peringkat Bertanda Wilcoxon merupakan uji statistik yang dilakukan untuk melihat apakah ada perbedaan median dari suatu observasi berpasangan dengan memperhitungkan besarnya selisih dari dua observasi yang bersesuaian. Uji Peringkat Bertanda Wilcoxon adalah suatu uji nonparametrik yang biasanya digunakan pada data kualitatif (skala nominal dan ordinal) atau untuk data kuantitatif yang tidak berdistribusi normal.

Perumusan hipotesis dalam uji peringkat bertanda Wilcoxon (Triola, 2009) adalah:

$$H_0 : Me_1 = Me_2.$$

$$H_1 : Me_1 \neq Me_2.$$

Cara pengujian:

- Tentukan selisih nilai pasangan yaitu d.
- Untuk nilai yang sama ($d=0$), data dieliminir
- Nilai d diranking (tanpa memperhatikan tanda + atau -)
- Nilai d yang sama, rankingnya adalah rata-ratanya
- Hitung T = jumlah ranking bertanda (+) atau (-) yang paling kecil

Statistik Uji:

Untuk $n \leq 30$: T

Untuk $n > 30$:

$$Z = \frac{T - \frac{n(n+1)}{4}}{\sqrt{\frac{n(n+1)(2n+1)}{24}}}$$

Kriteria:

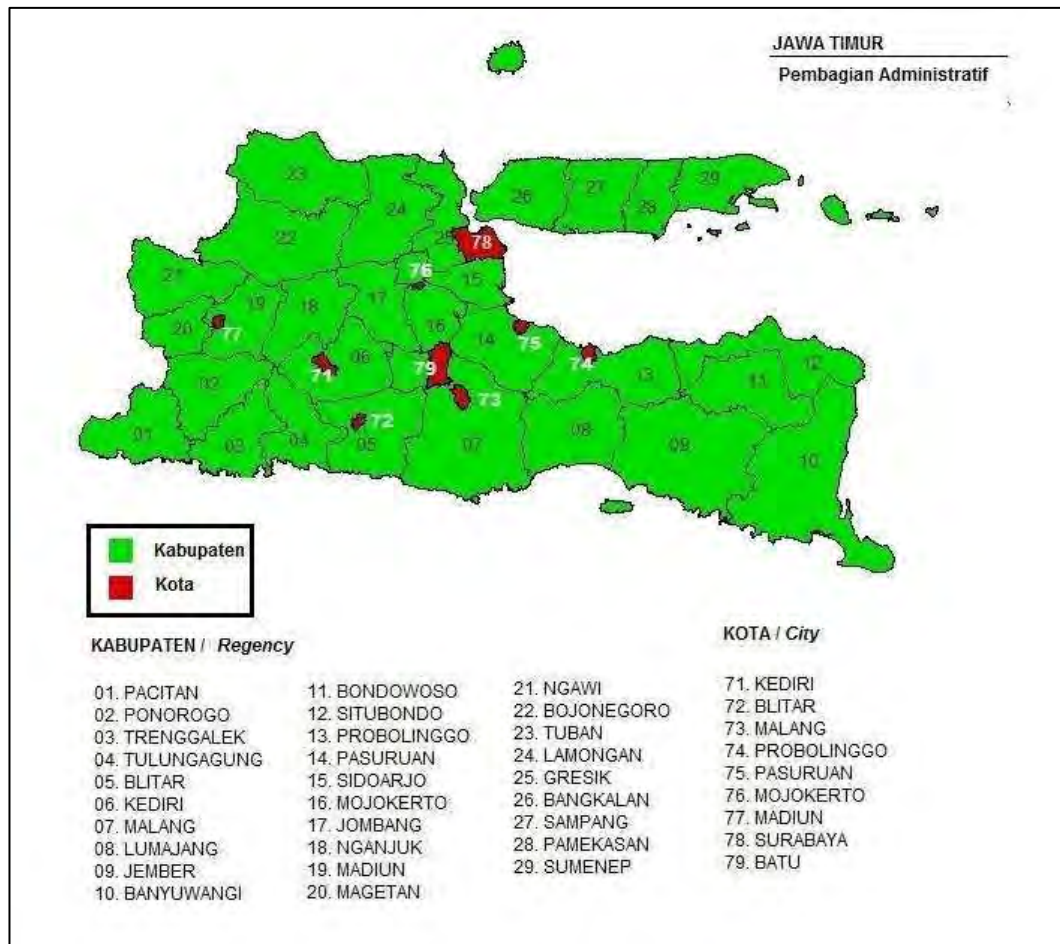
Untuk $n \leq 30$: Tolak H_0 jika $T \leq T_\alpha$.

Untuk $n > 30$: Tolak H_0 jika $|Z| > Z_{\alpha/2}$.

2.7 Gambaran Umum Provinsi Jawa Timur

Provinsi Jawa Timur merupakan salah satu provinsi di Indonesia. Secara geografis, Jawa Timur terletak di antara $111^\circ 0'$ Bujur Timur – $114^\circ 4'$ Bujur Timur dan $7^\circ 12'$ Lintang Selatan – $8^\circ 48'$ Lintang Selatan. Jawa Timur berbatasan dengan Provinsi Kalimantan Selatan di sebelah utara, Pulau Bali di sebelah timur,

Samudera Hindia di sebelah selatan dan Provinsi Jawa Tengah di sebelah Barat. Jawa Timur memiliki wilayah seluas 47.995 km² yang meliputi dua bagian utama yaitu Jawa Timur daratan dan Kepulauan Madura. Luas wilayah Jawa Timur daratan hampir mencakup 90 persen dari seluruh luas wilayah Provinsi Jawa Timur, sedangkan luas Kepulauan Madura hanya sekitar 10 persen.



Gambar 2.2 Peta Wilayah Administrasi Provinsi Jawa Timur.
(Sumber : BPS Provinsi Jawa Timur)

Wilayah Jawa Timur berada di sekitar garis khatulistiwa, sehingga wilayah ini memiliki dua perubahan musim setiap tahunnya, yaitu musim kemarau yang terjadi pada bulan Juni – Oktober dan musim penghujan pada bulan Nopember – Mei. Provinsi Jawa Timur dapat dibedakan menjadi tiga dataran yaitu dataran tinggi, sedang dan rendah. Dataran tinggi merupakan daerah dengan ketinggian rata-rata di atas 100 meter di atas permukaan laut. Daerah ini meliputi Kabupaten Magetan, Kabupaten Trenggalek, Kabupaten Blitar, Kabupaten Malang,

Kabupaten Bondowoso, Kota Blitar, Kota Malang, dan Kota Batu. Dataran sedang mempunyai ketinggian antara 45 - 100 meter di atas permukaan laut. Daerah ini meliputi Kabupaten Tulungagung, Kediri, Lumajang, Jember, Nganjuk, Madiun, Ponorogo, Ngawi, Bangkalan, Kota Kediri dan Kota Madiun. Sedangkan kabupaten dan kota lainnya merupakan dataran rendah, dengan ketinggian rata-rata 45 meter dari permukaan laut yang terdiri dari 15 kabupaten dan 4 kota.

Wilayah Jawa Timur berada di sekitar garis khatulistiwa, sehingga wilayah ini memiliki dua perubahan musim setiap tahunnya, yaitu musim kemarau yang terjadi pada bulan Juni – Oktober dan musim penghujan pada bulan Nopember – Mei. Provinsi Jawa Timur dapat dibedakan menjadi tiga dataran yaitu dataran tinggi, sedang dan rendah. Dataran tinggi merupakan daerah dengan ketinggian rata-rata di atas 100 meter di atas permukaan laut. Daerah ini meliputi Kabupaten Magetan, Kabupaten Trenggalek, Kabupaten Blitar, Kabupaten Malang, Kabupaten Bondowoso, Kota Blitar, Kota Malang, dan Kota Batu. Dataran sedang mempunyai ketinggian antara 45 - 100 meter di atas permukaan laut. Daerah ini meliputi Kabupaten Tulungagung, Kediri, Lumajang, Jember, Nganjuk, Madiun, Ponorogo, Ngawi, Bangkalan, Kota Kediri dan Kota Madiun. Sedangkan kabupaten dan kota lainnya merupakan dataran rendah, dengan ketinggian rata-rata 45 meter dari permukaan laut yang terdiri dari 15 kabupaten dan 4 kota.

Jawa Timur merupakan provinsi yang memiliki beberapa gunung berapi yaitu Gunung Kelud, Gunung Welirang, Gunung Arjuno, Gunung Semeru, Gunung Lamongan, Gunung Bromo, Gunung Argopuro, Gunung Pendil, Gunung Suket, Gunung Merapi, dan Gunung Raung. Selain itu, Jawa Timur juga memiliki sungai-sungai besar dan kecil yang melintasi berbagai wilayah kabupaten/kota. Kondisi alam tersebut sangat mendukung kegiatan pertanian yang banyak diusahakan oleh masyarakat di Jawa Timur.

Secara administratif, Jawa Timur terbagi menjadi 29 kabupaten dan 9 kota, dengan Kota Surabaya sebagai ibukota provinsi. Hal ini menjadikan Jawa Timur sebagai provinsi dengan jumlah kabupaten/kota terbanyak di Indonesia. Jumlah penduduk Jawa Timur pada tahun 2014 mencapai 38.610.202 jiwa dengan laju pertumbuhan penduduk sebesar 0,64 persen/tahun (BPS, 2015). Kota Surabaya mempunyai jumlah penduduk yang paling besar, yaitu 2.833.924 jiwa, diikuti

Kabupaten Malang 2.527.087 jiwa dan Kabupaten Jember 2.394.608 jiwa. Kepadatan penduduk Jawa Timur tahun 2014 adalah 804 jiwa/km². Secara umum, kepadatan penduduk di kota lebih tinggi dibanding dengan kepadatan penduduk di kabupaten. Kota Surabaya mempunyai kepadatan penduduk tertinggi yaitu 8.562 jiwa/km². Hal ini disebabkan oleh daya tarik Kota Surabaya yang merupakan salah satu kota favorit di Jawa Timur dengan berbagai ketersediaan fasilitas umum dan lapangan kerja yang relatif lebih beragam.

Tabel 2.2 Jumlah, Laju Pertumbuhan dan Kepadatan Penduduk menurut Kabupaten/Kota di Jawa Timur Tahun 2014

No	Kabupaten/Kota	Jumlah penduduk (jiwa)	Laju pertumbuhan penduduk (%)	Kepadatan penduduk (jiwa/km ²)
1.	Pacitan	549 481	0,29	387
2.	Ponorogo	865 809	0,22	612
3.	Trenggalek	686 781	0,44	552
4.	Tulungagung	1 015 974	0,65	883
5.	Blitar	1 140 793	0,36	651
6.	Kediri	1 538 929	0,55	1 011
7.	Malang	2 527 087	0,73	731
8.	Lumajang	1 026 378	0,25	569
9.	Jember	2 394 608	0,55	722
10.	Banyuwangi	1 588 082	0,35	442
11.	Bondowoso	756 989	0,56	482
12.	Situbondo	666 013	0,80	403
13.	Probolinggo	1 132 690	0,84	664
14.	Pasuruan	1 569 507	0,82	1 056
15.	Sidoarjo	2 083 924	1,71	2 898
16.	Mojokerto	1 070 486	1,20	1 099
17.	Jombang	1 234 501	0,29	1 108
18.	Nganjuk	1 037 723	0,40	808
19.	Madiun	673 988	0,31	602
20.	Magetan	626 614	0,15	888
21.	Ngawi	827 829	0,39	594
22.	Bojonegoro	1 232 386	0,38	532
23.	Tuban	1 147 097	0,49	580
24.	Lamongan	1 187 084	0,06	675

Lanjutan Tabel 2.2

No	Kabupaten/Kota	Jumlah penduduk (jiwa)	Laju pertumbuhan penduduk (%)	Kepadatan penduduk (jiwa/km ²)
25.	Gresik	1 241 613	1,18	1 003
26.	Bangkalan	945 821	0,89	726
27.	Sampang	925 911	1,36	750
28.	Pamekasan	836 224	1,07	1 051
29.	Sumenep	1 067 202	0,56	512
30.	Kota Kediri	278 072	0,53	4 030
31.	Kota Blitar	136 903	0,89	4 149
32.	Kota Malang	845 973	0,61	7 691
33.	Kota Probolinggo	226 777	1,29	4 200
34.	Kota Pasuruan	193 329	0,54	5 088
35.	Kota Mojokerto	124 719	0,74	6 236
36.	Kota Madiun	174 373	0,15	5 129
37.	Kota Surabaya	2 833 924	0,43	8 562
38.	Kota Batu	198 608	1,23	983
Jawa Timur		38 610 202	0,64	804

Sumber: BPS

2.8 Desa Tertinggal

Desa adalah desa dan desa adat atau yang disebut dengan nama lain, selanjutnya disebut desa adalah kesatuan masyarakat hukum yang memiliki batas wilayah yang berwenang untuk mengatur dan mengurus urusan pemerintahan, kepentingan masyarakat setempat berdasarkan prakarsa masyarakat, hak asal usul, dan/atau hak tradisional yang diakui dan dihormati dalam sistem pemerintahan Negara Kesatuan Republik Indonesia (Undang-Undang Nomor 6 Tahun 2014 tentang Desa).

Desa tertinggal adalah desa-desa yang kondisinya secara ekonomi relatif tertinggal dibandingkan desa-desa lainnya. Kemajuan atau ketertinggalan suatu desa dicerminkan oleh indikator utama, yaitu tinggi rendahnya rata-rata pengeluaran per kapita penduduk desa. Beberapa faktor diduga menjadi penyebab kemajuan atau ketertinggalan suatu desa, yaitu faktor alam/lingkungan, faktor kelembagaan, faktor sarana/prasarana dan akses, serta faktor sosial ekonomi penduduk (BPS, 2005).

Beberapa faktor yang diduga menjadi penyebab kemajuan atau ketertinggalan suatu desa yaitu perekonomian masyarakat, sumber daya manusia, sarana dan prasarana (infrastruktur), kemampuan keuangan lokal (celah fiskal), aksesibilitas, dan karakteristik daerah (Edy, 2009).

Menurut kemendagri (2012), desa tertinggal adalah desa/kelurahan/nagari/ UPT/ nama komunitas hukum lainnya yang potensi umumnya rendah, laju perkembangannya lamban dan kurang berkembang serta status perkembangannya berada pada tingkat swadaya dengan kategori mula, madya dan lanjut.

Pada tahun 2015, Badan Perencanaan Nasional dan Badan Pusat Statistik menerbitkan Indeks Pembangunan Desa 2014 yang memuat klasifikasi desa. Berdasarkan IPD tahun 2014, desa dibagi menjadi tiga klasifikasi yaitu Desa Mandiri, Desa Berkembang, dan Desa Tertinggal. Desa mandiri berjumlah 2.898 desa, desa berkembang berjumlah 50.763 desa, dan desa tertinggal berjumlah 20.432 desa. Desa Mandiri adalah desa yang telah terpenuhi SPM desa mencakup beberapa aspek yaitu kebutuhan sosial dasar, infrastruktur dasar, sarana dasar, pelayanan umum, dan penyelenggaraan pemerintahan desa, serta kelembagaan desa yang keberlanjutan. Desa Berkembang adalah desa yang sudah terpenuhi SPM desa pada semua aspek tetapi pengelolaannya belum menunjukkan keberlanjutan. Sedangkan, Desa Tertinggal adalah desa yang belum terpenuhi SPM desa pada aspek kebutuhan sosial dasar, infrastruktur dasar, sarana dasar, pelayanan umum, dan penyelenggaraan pemerintahan.

IPD terdiri dari lima dimensi yang disesuaikan dengan ketersediaan data/variabel dalam data Potensi Desa 2014 yaitu sebagai berikut (Bappenas dan BPS, 2015):

1. **Pelayanan Dasar** mewakili aspek pelayanan dasar untuk mewujudkan bagian dari kebutuhan dasar, khusus untuk pendidikan dan kesehatan. Variabel yang termasuk sebagai komponen penyusunnya meliputi ketersediaan dan akses terhadap fasilitas pendidikan seperti TK, SD, SMP, dan SMA; serta ketersediaan dan akses terhadap fasilitas kesehatan seperti rumah sakit, rumah sakit bersalin, puskesmas/pustu, tempat praktek dokter, poliklinik/balai pengobatan, tempat praktek bidan, poskesdes, polindes, dan apotek.

2. **Kondisi Infrastruktur** mewakili Kebutuhan Dasar; Sarana; Prasarana; Pengembangan Ekonomi Lokal; dan Pemanfaatan Sumberdaya Alam secara Berkelanjutan dengan memisahkan aspek aksesibilitas/transportasi. Variabel-variabel penyusunnya mencakup ketersediaan infrastruktur ekonomi seperti: kelompok pertokoan, minimarket, maupun toko kelontong, pasar, restoran, rumah makan, maupun warung/kedai makanan, akomodasi hotel atau penginapan, serta bank; ketersediaan infrastruktur energi seperti: listrik, penerangan jalan, dan bahan bakar untuk memasak; ketersediaan infrastruktur air bersih dan sanitasi seperti: sumber air minum, sumber air mandi/cuci, dan fasilitas buang air besar; serta ketersediaan dan kualitas infrastruktur komunikasi dan informasi seperti: komunikasi menggunakan telepon seluler, internet, dan pengiriman pos/barang.
3. **Aksesibilitas/Transportasi** dipisahkan sebagai dimensi tersendiri dalam indikator pembangunan desa dengan pertimbangan sarana dan prasarana transportasi memiliki kekhususan dan prioritas pembangunan desa sebagai penghubung kegiatan sosial ekonomi dalam desa. Variabel-variabel penyusunnya meliputi ketersediaan dan akses terhadap sarana transportasi seperti: lalu lintas dan kualitas jalan, aksesibilitas jalan, ketersediaan dan operasional angkutan umum; dan aksesibilitas transportasi seperti: waktu tempuh per kilometer transportasi ke kantor camat, biaya per kilometer transportasi ke kantor camat, waktu tempuh per kilometer transportasi ke kantor bupati/walikota, dan biaya per kilometer transportasi ke kantor bupati/walikota.

Indikator jarak antara desa dengan pusat pemerintahan ini merujuk dari variabel yang tercantum dalam Potensi Desa 2014, dengan asumsi bahwa pada umumnya pusat kegiatan sosial-ekonomi suatu kawasan berada di sekitar/dekat pusat-pusat pemerintahan.
4. **Pelayanan Umum** merupakan upaya pemenuhan kebutuhan pelayanan atas barang, jasa, dan/atau pelayanan administratif dengan tujuan memperkuat demokrasi, kohesi sosial, perlindungan lingkungan, dan sebagainya. Karena kekhususannya, variabel pelayanan administratif dinyatakan sebagai dimensi tersendiri (Penyelenggaraan Pemerintahan). Begitupun dengan variabel

pendidikan, kesehatan, transportasi, dan lainnya menjadi dimensi tersendiri yang telah dijelaskan sebelumnya. Pelayanan dalam dimensi ini mewakili aspek lingkungan dan aspek pemberdayaan masyarakat serta mengacu pada ketersediaan data Potensi Desa 2014. Aspek lingkungan dalam hal ini terkait dengan kesehatan lingkungan masyarakat, sedangkan aspek pemberdayaan masyarakat diwakili dengan keberadaan kelompok kegiatan masyarakat. Oleh karena itu, variabel-variabel penyusun dimensi ini mencakup penanganan kesehatan masyarakat seperti: penanganan kejadian luar biasa (KLB), dan penanganan gizi buruk; serta ketersediaan fasilitas olah raga seperti: ketersediaan lapangan olah raga, dan kelompok kegiatan olah raga.

5. **Penyelenggaraan Pemerintahan** mewakili indikasi kinerja pemerintahan desa merupakan bentuk pelayanan administratif yang diselenggarakan penyelenggara pelayanan bagi warga yang dalam hal ini adalah Pemerintah. Oleh karena itu variabel ini perlu diukur dan berdiri sendiri sebagai sebuah indikator pembangunan desa, karena sifatnya sebagai perangkat terlaksananya tujuan pembangunan desa tersebut. Variabel-variabel penyusunnya meliputi kemandirian seperti: kelengkapan pemerintahan desa, otonomi desa, dan asset/kekayaan desa; serta kualitas sumber daya manusia seperti: kualitas SDM kepala desa dan sekretaris desa.

IPD merupakan indeks komposit tertimbang dari 42 indikator yang secara substansi dan bersama-sama menggambarkan tingkat pembangunan di desa. Setiap indikator harus memiliki kontribusi terhadap IPD. Besarnya kontribusi setiap indikator menggambarkan besarnya pengaruh indikator tersebut terhadap dimensi dan IPD. Besarnya kontribusi setiap indikator diperoleh dengan menggunakan metode Analisis Komponen Utama (*Principal Component Analysis*).

Nilai IPD diperoleh dari penjumlahan secara tertimbang terhadap setiap Indikator penyusun IPD. Nilai yang dijumlahkan adalah skor setiap indikator yang sudah ditimbang/dikalikan dengan penimbang masing-masing indikator. Nilai indeks mempunyai rentang 0 s/d 100. Desa mandiri merupakan desa dengan nilai IPD lebih dari 75. Desa berkembang merupakan desa yang memiliki

nilai IPD lebih dari 50 namun kurang dari atau sama dengan 75. Desa tertinggal merupakan desa yang memiliki nilai IPD kurang dari atau sama dengan 50.

BAB 3

METODOLOGI PENELITIAN

Bab ini terdiri dari dua sub bab yaitu metodologi tujuan penelitian pertama untuk menaksir parameter *Rare Event Weighted Logistic Regression* dan metodologi tujuan penelitian kedua dan ketiga untuk penerapan dan perbandingan *Rare Event Weighted Logistic Regression* pada klasifikasi *imbalanced data*. Studi kasus yang digunakan yaitu klasifikasi desa tertinggal di provinsi Jawa Timur.

3.1 Penaksiran Parameter *Rare Event Weighted Logistic Regression*

Langkah-langkah untuk memperoleh bentuk penaksir parameter pada model *RE-WLR*

a. *Regularized Logistic Regression*

- 1) Menentukan fungsi kemungkinan (*likelihood function*) pada model *Logistic Regression*.
- 2) Membentuk *regularized log-likelihood function* yang diperoleh dengan menentukan *log-likelihood function* pada langkah (a.1) dan menambahkan parameter regularisasi $\frac{\lambda}{2} \|\beta\|^2$.

b. *Regularized Weighted Logistic Regression*

- 1) Menentukan pembobot.
- 2) Menentukan fungsi kemungkinan (*likelihood function*) pada model *Weighted Logistic Regression*.
- 3) Membentuk *regularized weighted log-likelihood function* $\ln L_w(\beta)$ yang diperoleh dengan menentukan *log-likelihood function* pada langkah (b.2) dan menambahkan parameter regularisasi $\frac{\lambda}{2} \|\beta\|^2$.
- 4) Mendapatkan vektor *gradient* dengan mencari turunan parsial pertama dari *regularized weighted log-likelihood function* (b.3), yaitu $\nabla_{\beta} \ln L_w(\beta)$.
- 5) Mendapatkan matriks *Hessian* dengan mencari turunan parsial kedua dari *regularized weighted log-likelihood function* (b.3), yaitu $\nabla_{\beta}^2 \ln L_w(\beta)$.

- 6) Mendapatkan iterasi Newton-Raphson untuk $\hat{\beta}$ menggunakan $\nabla_{\beta} \ln L_w(\beta)$ dan $\nabla_{\beta}^2 \ln L_w(\beta)$ dari langkah (b.4) dan (b.5).
- 7) Mendapatkan penaksir parameter $\hat{\beta}$ dengan menggunakan metode Truncated Newton untuk subproblem *WLS* dengan algoritma *CG* linier.
- 8) Menentukan vektor bias $B(\hat{\beta})$.
- 9) Mendapatkan penaksir parameter $\tilde{\beta}$ dengan bias terkoreksi.

3.2 Penerapan *Rare Event Weighted Logistic Regression*

3.2.1 Sumber Data

Data yang digunakan dalam penelitian ini adalah data Potensi Desa (PODES) Provinsi Jawa Timur Tahun 2014 serta data Indeks Pembangunan Desa 2014 yang dikeluarkan oleh Bappenas dan Badan Pusat Statistik (BPS).

Pendataan PODES dilaksanakan setiap tiga tahun sekali oleh BPS. Secara umum tujuan Podes 2014 adalah:

1. Menyediakan data yang diharapkan dapat mendukung perencanaan kegiatan Sensus Ekonomi 2016 dari sisi wilayah kerja, anggaran, dan alokasi petugas,
2. Sebagai sarana untuk *updating* Master File Desa (MFD),
3. Menyediakan data tentang keberadaan dan perkembangan potensi yang dimiliki desa/kelurahan yang meliputi: sosial, ekonomi, sarana, dan prasarana wilayah,
4. Menyediakan data untuk berbagai keperluan yang berkaitan dengan perencanaan wilayah di tingkat nasional dan tingkat daerah,
5. Melengkapi penyusunan kerangka sampling (*sampling frame*) untuk kegiatan statistik lain lebih lanjut,
6. Menyediakan data bagi keperluan *updating* klasifikasi/tipologi desa, misalnya perkotaan-perdesaan, pesisir dan nonpesisir, dan sebagainya,
7. Menyediakan data bagi keperluan *updating* peta wilayah kerja statistik terendah,

8. Menyediakan data pokok bagi penyusunan statistik wilayah kecil (*small area statistics*),
9. Menyediakan data bagi penyusunan berbagai analisis seperti identifikasi dan penentuan desa tertinggal, variabel konteks dalam PMT, dan identifikasi desa rawan bencana.

3.2.2 Variabel Penelitian

Variabel-variabel dalam penelitian ini dipilih berdasarkan referensi dari publikasi Indeks Pembangunan Desa 2014 yang memuat variabel-variabel yang digunakan oleh Bappenas dan BPS dalam penentuan klasifikasi desa. Penelitian dilakukan pada seluruh desa yang ada di Provinsi Jawa Timur.

Variabel respon (Y) merupakan variabel yang berisi kelas yang terdiri atas dua kategori yaitu {0} untuk desa tidak tertinggal dan {1} untuk desa tertinggal. Pengukuran variabel respon didapat dari Indeks Pembangunan Desa 2014 yang mengklasifikasikan desa menjadi desa tertinggal, desa berkembang, dan desa mandiri. Dalam penelitian ini, klasifikasi dibagi menjadi dua kategori (desa tertinggal dan desa tidak tertinggal), sehingga desa berkembang dan desa mandiri dikategorikan ke dalam desa tidak tertinggal. Data variabel prediktor didapatkan dari hasil pendataan PODES Provinsi Jawa Timur Tahun 2014.

Tabel 3.1 Variabel dalam Penelitian

Variabel	Keterangan
Y	Status ketertinggalan desa
X ₁	Ketersediaan TK/RA/BA
X ₂	Ketersediaan SD Sederajat
X ₃	Ketersediaan SMP Sederajat
X ₄	Ketersediaan SMA Sederajat
X ₅	Ketersediaan Rumah Sakit
X ₆	Ketersediaan Rumah Sakit Bersalin
X ₇	Ketersediaan Puskesmas
X ₈	Ketersediaan Poliklinik/Balai Pengobatan
X ₉	Ketersediaan Tempat Praktek Dokter
X ₁₀	Ketersediaan Tempat Praktek Bidan
X ₁₁	Ketersediaan Poskesdes atau Polindes
X ₁₂	Ketersediaan Apotek
X ₁₃	Ketersediaan Pertokoan, Minimarket atau Toko Kelontong
X ₁₄	Ketersediaan Pasar
X ₁₅	Ketersediaan Restoran, Rumah Makan atau Warung Makan

Lanjutan Tabel 3.1

Variabel	Keterangan
X ₁₆	Ketersediaan Akomodasi Hotel atau Penginapan
X ₁₇	Ketersediaan Bank
X ₁₈	Elektrifikasi
X ₁₉	Kondisi Penerangan di Jalan Utama
X ₂₀	Bahan Bakar Untuk Memasak
X ₂₁	Sumber Air Untuk Minum
X ₂₂	Sumber Air Untuk Mandi/Cuci
X ₂₃	Fasilitas Buang Air Besar
X ₂₄	Kualitas Fasilitas Komunikasi Seluler
X ₂₅	Ketersediaan Fasilitas Internet dan Pengiriman Pos atau Barang
X ₂₆	Lalu Lintas dan Kualitas Jalan
X ₂₇	Aksesibilitas Jalan
X ₂₈	Ketersediaan Angkutan Umum
X ₂₉	Operasional Angkutan Umum
X ₃₀	Waktu Tempuh per Kilometer Transportasi ke Kantor Camat
X ₃₁	Biaya per Kilometer Transportasi ke Kantor Camat
X ₃₂	Waktu Tempuh per Kilometer Transportasi ke Kantor Bupati/Walikota
X ₃₃	Biaya per Kilometer Transportasi ke Kantor Bupati/Walikota
X ₃₄	Penanganan Kejadian Luar Biasa (KLB)
X ₃₅	Penanganan Gizi Buruk
X ₃₆	Ketersediaan Fasilitas Olah Raga
X ₃₇	Keberadaan Kelompok Kegiatan Olah Raga
X ₃₈	Kelengkapan Pemerintahan Desa
X ₃₉	Otonomi Desa
X ₄₀	Aset/Kekayaan Desa
X ₄₁	Kualitas SDM Kepala Desa
X ₄₂	Kualitas SDM Sekretaris Desa

Struktur data untuk penelitian ini ditunjukkan pada Tabel 3.2.

Tabel 3.2 Struktur Data dalam Penelitian

Desa	Y	X ₁	X ₂	X ₃	X ₄	X ₅	...	X ₄₂
1	y ₁	x _{1.1}	x _{1.2}	x _{1.3}	x _{1.4}	x _{1.5}	...	x _{1.42}
2	y ₂	x _{2.1}	x _{2.2}	x _{2.3}	x _{2.4}	x _{2.5}	...	x _{2.42}
3	y ₃	x _{3.1}	x _{3.2}	x _{3.3}	x _{3.4}	x _{3.5}	...	x _{3.42}
4	y ₄	x _{4.1}	x _{4.2}	x _{4.3}	x _{4.4}	x _{4.5}	...	x _{4.42}
5	y ₅	x _{5.1}	x _{5.2}	x _{5.3}	x _{5.4}	x _{5.5}	...	x _{5.42}
⋮	⋮	⋮	⋮	⋮	⋮	⋮	...	⋮
7721	y ₇₇₂₁	x _{7721.1}	x _{7721.2}	x _{7721.3}	x _{7721.4}	x _{7721.5}	...	x _{7721.42}

Definisi operasional variabel penelitian sebagai berikut:

1. Status ketertinggalan desa. Desa tertinggal adalah desa-desa yang kondisinya relatif tertinggal dibandingkan desa-desa lainnya.

Kategori dari Bappenas:

- 1) Desa Tertinggal
- 2) Desa Berkembang
- 3) Desa Mandiri

Rekategorisasi:

- 1) Desa Tertinggal
- 2) Desa Tidak Tertinggal

2. TK/RA/BA, meliputi Taman Kanak-Kanak (TK), Bustanuf Athfal (BA) dan Raudatul Athfal (RA).

Kategori dari Bappenas:

- 1) **Kode 0 (nol)** merupakan kondisi dimana desa tersebut tidak terdapat TK/RA/BA. Jarak menuju TK/RA/BA terdekat sejauh lebih dari atau sama dengan 20 km.
- 2) **Kode 1 (satu)** merupakan kondisi dimana desa tersebut tidak terdapat TK/RA/BA. Jarak menuju TK/RA/BA terdekat sejauh lebih dari atau sama dengan 10 km dan kurang dari 20 km.
- 3) **Kode 2 (dua)** merupakan kondisi dimana desa tersebut tidak terdapat TK/RA/BA. Jarak menuju TK/RA/BA kurang dari 10 km.
- 4) **Kode 3 (tiga)** merupakan kondisi dimana di desa tersebut terdapat TK/RA/BA dengan rasio ketersediaan TK/RA/BA terhadap penduduk desa kurang dari 0.00047665382434 (5 fasilitas per 10.000 penduduk).
- 5) **Kode 4 (empat)** merupakan kondisi dimana di desa tersebut terdapat TK/RA/BA dengan rasio ketersediaan TK/RA/BA terhadap penduduk desa lebih dari atau sama dengan 0.00047665382434 (5 fasilitas per 10.000 penduduk) dan kurang dari 0.00095330764868 (10 fasilitas per 10.000 penduduk).
- 6) **Kode 5 (lima)** merupakan kondisi dimana di desa tersebut terdapat TK/RA/BA dengan rasio ketersediaan TK/RA/BA terhadap penduduk

desa lebih dari atau sama dengan 0.00095330764868 (10 fasilitas per 10.000 penduduk).

Rekategorisasi:

- 1) Tidak tersedia TK/RA/BA.
- 2) Tersedia TK/RA/BA.
3. SD/MI, meliputi Sekolah Dasar, Madrasah Ibtidaiyah (MI), baik negeri maupun swasta

Kategori dari Bappenas:

- 1) **Kode 0 (nol)** merupakan kondisi dimana desa tersebut tidak terdapat SD sederajat. Jarak menuju SD sederajat terdekat sejauh lebih dari atau sama dengan 8 km.
- 2) **Kode 1 (satu)** merupakan kondisi dimana desa tersebut tidak terdapat SD sederajat. Jarak menuju SD sederajat terdekat sejauh lebih dari atau sama dengan 4 km dan kurang dari 8 km.
- 3) **Kode 2 (dua)** merupakan kondisi dimana desa tersebut tidak terdapat SD sederajat. Jarak menuju SD sederajat kurang dari 4 km.
- 4) **Kode 3 (tiga)** merupakan kondisi dimana di desa tersebut terdapat SD sederajat dengan rasio ketersediaan SD sederajat terhadap penduduk desa kurang dari 0.00069840871109 (7 fasilitas per 10.000 penduduk).
- 5) **Kode 4 (empat)** merupakan kondisi dimana di desa tersebut terdapat SD sederajat dengan rasio ketersediaan SD sederajat terhadap penduduk desa lebih dari atau sama dengan 0,00069840871109 (7 fasilitas per 10.000 penduduk) dan kurang dari 0,00139681742219 (14 fasilitas per 10.000 penduduk).
- 6) **Kode 5 (lima)** merupakan kondisi dimana di desa tersebut terdapat SD sederajat dengan rasio ketersediaan SD sederajat terhadap penduduk desa lebih dari atau sama dengan 0,00139681742219 (14 fasilitas per 10.000 penduduk).

Rekategorisasi:

- 1) Tidak tersedia SD.
- 2) Tersedia SD.

4. SMP sederajat, meliputi Sekolah Menengah Pertama, Madrasah Tsanawiyah (MTs), baik negeri maupun swasta.

Kategori dari Bappenas:

- 1) **Kode 0 (nol)** merupakan kondisi dimana desa tersebut tidak terdapat SMP sederajat. Jarak menuju SMP sederajat terdekat sejauh lebih dari atau sama dengan 9 km.
- 2) **Kode 1 (satu)** merupakan kondisi dimana desa tersebut tidak terdapat SMP sederajat. Jarak menuju SMP sederajat terdekat sejauh lebih dari atau sama dengan 6 km dan kurang dari 9 km.
- 3) **Kode 2 (dua)** merupakan kondisi dimana desa tersebut tidak terdapat SMP sederajat. Jarak menuju SMP sederajat terdekat lebih dari atau sama dengan 3 km dan kurang dari 6 km.
- 4) **Kode 3 (tiga)** merupakan kondisi dimana di desa tersebut tidak terdapat SMP sederajat. Jarak menuju SMP sederajat terdekat kurang dari 3 km.
- 5) **Kode 4 (empat)** merupakan kondisi dimana di desa tersebut terdapat SMP sederajat dengan rasio ketersediaan SMP sederajat terhadap penduduk desa kurang dari 0,00053531809209 (5 fasilitas per 10.000 penduduk).
- 6) **Kode 5 (lima)** merupakan kondisi dimana di desa tersebut terdapat SMP sederajat dengan rasio ketersediaan SMP sederajat terhadap penduduk desa lebih dari atau sama dengan 0,00053531809209 (5 fasilitas per 10.000 penduduk).

Rekategorisasi:

- 1) Tidak tersedia SMP.
 - 2) Tersedia SMP.
5. SMU sederajat, meliputi Sekolah Menengah Umum, Madrasah Aliyah (MA), Sekolah Menengah Kejuruan, Madrasah Aliyah Kejuruan (MAK), baik negeri maupun swasta.

Kategori dari Bappenas:

- 1) **Kode 0 (nol)** merupakan kondisi dimana desa tersebut tidak terdapat SMA sederajat. Jarak menuju SMA sederajat terdekat sejauh lebih dari atau sama dengan 14 km.

- 2) **Kode 1 (satu)** merupakan kondisi dimana desa tersebut tidak terdapat SMA sederajat. Jarak menuju SMA sederajat terdekat sejauh lebih dari atau sama dengan 9 km dan kurang dari 14 km.
- 3) **Kode 2 (dua)** merupakan kondisi dimana desa tersebut tidak terdapat SMA sederajat. Jarak menuju SMA sederajat terdekat lebih dari atau sama dengan 5 km dan kurang dari 9 km.
- 4) **Kode 3 (tiga)** merupakan kondisi dimana di desa tersebut tidak terdapat SMA sederajat. Jarak menuju SMA sederajat terdekat kurang dari 5 km.
- 5) **Kode 4 (empat)** merupakan kondisi dimana di desa tersebut terdapat SMA sederajat dengan rasio ketersediaan SMA sederajat terhadap penduduk desa kurang dari 0,00044061849597 (4 fasilitas per 10.000 penduduk).
- 6) **Kode 5 (lima)** merupakan kondisi dimana di desa tersebut terdapat SMA sederajat dengan rasio ketersediaan SMA sederajat terhadap penduduk desa lebih dari atau sama dengan 0,00044061849597 (4 fasilitas per 10.000 penduduk).

Rekategorisasi:

- 1) Tidak tersedia SMU.
 - 2) Tersedia SMU.
6. Rumah Sakit (RS) adalah sarana kesehatan/bangunan tempat untuk melayani penderita yang sakit untuk berobat rawat jalan atau rawat inap yang pelayanannya disediakan oleh dokter, perawat dan tenaga ahli kesehatan lainnya. Rumah sakit yang dicatat adalah rumah sakit umum dan khusus.

Kategori dari Bappenas:

- 1) **Kode 0 (nol)** merupakan desa yang tidak terdapat rumah sakit. Jarak menuju rumah sakit terdekat sejauh lebih dari atau sama dengan 23 km dan untuk mencapai fasilitas tersebut dirasakan sulit atau sangat sulit.
- 2) **Kode 1 (satu)** merupakan kondisi desa yang tidak terdapat rumah sakit. Jarak menuju rumah sakit terdekat sejauh lebih dari atau sama dengan 23 km dan untuk mencapai fasilitas tersebut dianggap mudah atau sangat mudah.

- 3) **Kode 2 (dua)** merupakan kondisi desa yang tidak terdapat rumah sakit. Jarak menuju rumah sakit terdekat kurang dari 23 km dan untuk mencapai dirasakan sulit atau sangat sulit.
- 4) **Kode 3 (tiga)** merupakan kondisi desa yang tidak terdapat rumah sakit. Jarak menuju rumah sakit terdekat kurang dari 23 km dan untuk mencapai fasilitas tersebut dianggap mudah atau sangat mudah.
- 5) **Kode 4 (empat)** merupakan desa yang terdapat rumah sakit dengan rasio ketersediaan rumah sakit terhadap penduduk desa kurang dari 0,00018983029742 (2 fasilitas per 10.000 penduduk).
- 6) **Kode 5 (lima)** merupakan desa yang terdapat rumah sakit dengan rasio ketersediaan rumah sakit terhadap penduduk desa lebih dari atau sama dengan 0,00018983029742 (2 fasilitas per 10.000 penduduk).

Rekategorisasi:

- 1) Tidak tersedia RS.
 - 2) Tersedia RS.
7. Rumah Sakit Bersalin adalah rumah sakit khusus untuk persalinan, dilengkapi pelayanan spesialis pemeriksaan kehamilan, persalinan, rawat inap dan rawat jalan ibu dan anak yang berada di bawah pengawasan dokter spesialis kandungan.

Rumah Bersalin adalah sarana pelayanan kesehatan dengan izin sebagai rumah bersalin, dilengkapi pelayanan pemeriksaan kehamilan, persalinan serta pemeriksaan ibu dan anak yang berada di bawah pengawasan bidan senior.

Kategori dari Bappenas:

- 1) **Kode 0 (nol)** merupakan desa yang tidak terdapat rumah sakit bersalin maupun rumah sakit. Jarak menuju rumah sakit bersalin terdekat lebih dari atau sama dengan 44 km, dan untuk mencapai fasilitas tersebut dirasakan sulit atau sangat sulit.
- 2) **Kode 1 (satu)** merupakan desa yang tidak terdapat rumah sakit bersalin maupun rumah sakit. Jarak menuju rumah sakit bersalin terdekat lebih dari atau sama dengan 44 km, dan untuk mencapai fasilitas tersebut dirasakan mudah atau sangat mudah.

- 3) **Kode 2 (dua)** merupakan desa yang tidak terdapat rumah sakit bersalin maupun rumah sakit. Jarak menuju rumah sakit bersalin terdekat kurang dari 44 km, dan untuk mencapai dirasakan sulit atau sangat sulit.
- 4) **Kode 3 (tiga)** merupakan desa yang tidak terdapat rumah sakit bersalin maupun rumah sakit. Jarak menuju rumah sakit bersalin terdekat kurang dari 44 km, dan untuk mencapai dirasakan mudah atau sangat mudah.
- Kode 4 (empat)** merupakan desa yang terdapat rumah sakit bersalin dengan rasio ketersediaan rumah sakit bersalin terhadap penduduk desa kurang dari 0,00017982346108 (2 fasilitas per 10.000 penduduk).
- 5) **Kode 5 (lima)** merupakan desa yang terdapat rumah sakit bersalin dengan rasio ketersediaan rumah sakit bersalin terhadap penduduk desa lebih dari atau sama dengan 0,00017982346108 (2 fasilitas per 10.000 penduduk). Termasuk jika di desa tersebut terdapat rumah sakit.

Rekategorisasi:

- 1) Tidak tersedia RS Bersalin.
 - 2) Tersedia RS Bersalin.
8. Puskesmas adalah sebagai unit pelayanan kesehatan milik pemerintah (pelaksana teknis dinas kesehatan kabupaten/kota) yang bertanggung jawab terhadap pelayanan kesehatan masyarakat untuk wilayah kecamatan, sebagian kecamatan, atau kelurahan/desa. Puskesmas memberikan pelayanan berobat jalan dengan rawat inap. Biasanya Puskesmas berada di setiap kecamatan dan dapat terdiri dari 2 – 3 puskesmas di dalam 1 kecamatan. Pada Podes 2014, Puskesmas dibedakan antara puskesmas yang menyediakan pelayanan rawat inap dan tidak.

Puskesmas Pembantu (Pustu) sebagai sarana kesehatan/bangunan yang dipakai sebagai pusat kesehatan masyarakat untuk wilayah yang lebih kecil, misal di desa/kelurahan. Pustu merupakan sarana kesehatan milik pemerintah yang berfungsi menunjang dan membantu memperluas jangkauan puskesmas dengan melaksanakan kegiatan-kegiatan yang dilakukan puskesmas dalam ruang lingkup wilayah yang lebih kecil serta jenis dan kompetensi pelayanan yang disesuaikan dengan kemampuan tenaga dan sarana yang tersedia. Pustu

memberikan pelayanan berobat jalan. Pustu bertanggung jawab ke puskesmas induk di kecamatan.

Kategori dari Bappenas:

- 1) **Kode 0 (nol)** merupakan desa yang tidak terdapat puskesmas/pustu, rumah sakit bersalin, maupun rumah sakit. Jarak menuju puskesmas/pustu terdekat lebih dari atau sama dengan 4 km dan untuk mencapai fasilitas tersebut dirasakan sulit atau sangat sulit.
- 2) **Kode 1 (satu)** merupakan desa yang tidak terdapat puskesmas/pustu, rumah sakit bersalin, maupun rumah sakit. Jarak menuju puskesmas/pustu terdekat lebih dari sama dengan 4 km, dan untuk mencapai fasilitas tersebut dirasakan mudah atau sangat mudah.
- 3) **Kode 2 (dua)** merupakan desa yang tidak terdapat puskesmas/pustu, rumah sakit bersalin, maupun rumah sakit. Jarak menuju puskesmas/pustu terdekat kurang dari 4 km, dan untuk mencapai fasilitas tersebut dirasakan sulit atau sangat sulit.
- 4) **Kode 3 (tiga)** merupakan desa yang tidak terdapat puskesmas/pustu, rumah sakit bersalin, maupun rumah sakit. Jarak menuju puskesmas/pustu terdekat kurang dari 4 km, dan untuk mencapai fasilitas tersebut dirasakan mudah atau sangat mudah.
- 5) **Kode 4 (empat)** merupakan desa yang terdapat puskesmas/pustu dengan rasio ketersediaan puskesmas/pustu terhadap penduduk desa kurang dari 0,00062957239928 (6 fasilitas per 10.000 penduduk). Termasuk jika di desa tersebut terdapat rumah sakit bersalin, tetapi tidak ada rumah sakit.
- 6) **Kode 5 (lima)** merupakan desa yang terdapat puskesmas/pustu dengan rasio ketersediaan puskesmas/pustu terhadap penduduk desa lebih dari atau sama dengan 0,00062957239928 (6 fasilitas per 10.000 penduduk). Termasuk jika di desa tersebut terdapat rumah sakit tanpa memperhatikan ketersediaan rumah sakit bersalin.

Rekategorisasi:

- 1) Tidak tersedia Puskesmas/Pustu.
- 2) Tersedia Puskesmas/Pustu.

9. Poliklinik adalah sarana kesehatan/bangunan yang dipakai untuk pelayanan berobat jalan. Biasanya dikelola oleh swasta atau organisasi keagamaan tertentu.

Balai pengobatan adalah tempat pemeriksaan kesehatan di bawah pengawasan menteri kesehatan.

Kategori dari Bappenas:

- 1) **Kode 0 (nol)** merupakan desa yang tidak terdapat poliklinik/balai pengobatan, puskesmas, pustu, rumah sakit bersalin, maupun rumah sakit. Jarak menuju poliklinik/balai pengobatan terdekat lebih dari atau sama dengan 19 km, dan untuk mencapai fasilitas tersebut dirasakan sulit atau sangat sulit.
- 2) **Kode 1 (satu)** merupakan desa yang tidak terdapat poliklinik/balai pengobatan, puskesmas, pustu, rumah sakit bersalin, maupun rumah sakit. Jarak menuju poliklinik/balai pengobatan terdekat lebih dari atau sama dengan 19 km, dan untuk mencapai fasilitas tersebut dirasakan mudah atau sangat mudah.
- 3) **Kode 2 (dua)** merupakan desa yang tidak terdapat poliklinik/balai pengobatan, puskesmas, pustu, rumah sakit bersalin, maupun rumah sakit. Jarak menuju poliklinik/balai pengobatan terdekat kurang dari 19 km, dan untuk mencapai fasilitas tersebut dirasakan sulit atau sangat sulit.
- 4) **Kode 3 (tiga)** merupakan desa yang tidak terdapat poliklinik/balai pengobatan, puskesmas, pustu, rumah sakit bersalin, maupun rumah sakit. Jarak menuju poliklinik/balai pengobatan terdekat kurang dari 19 km, dan untuk mencapai fasilitas tersebut dirasakan mudah atau sangat mudah.
- 5) **Kode 4 (empat)** merupakan desa yang terdapat poliklinik/balai pengobatan dengan rasio ketersediaan poliklinik/balai pengobatan terhadap penduduk desa kurang dari 0,00028457719787 (3 fasilitas per 10.000 penduduk). Termasuk jika di desa tersebut terdapat puskesmas/pustu, tetapi tidak ada rumah sakit bersalin maupun rumah sakit.

- 6) **Kode 5 (lima)** merupakan desa yang terdapat poliklinik/balai pengobatan dengan rasio ketersediaan poliklinik/balai pengobatan terhadap penduduk desa lebih dari atau sama dengan 0,00028457719787 (3 fasilitas per 10.000 penduduk). Termasuk jika di desa tersebut terdapat rumah sakit bersalin atau rumah sakit, tanpa mempertimbangkan ketersediaan puskesmas/pustu.

Rekategorisasi:

- 1) Tidak tersedia poliklinik/balai pengobatan.
 - 2) Tersedia poliklinik/balai pengobatan.
10. Tempat praktek dokter adalah sarana kesehatan/bangunan yang digunakan untuk tempat praktek dokter yang biasanya memberikan pelayanan berobat jalan, termasuk praktek dokter yang mempunyai fasilitas rawat inap dan apotek.

Kategori dari Bappenas:

- 1) **Kode 0 (nol)** merupakan desa yang tidak terdapat tempat praktek dokter, poliklinik/balai pengobatan, puskesmas, pustu, rumah sakit bersalin, maupun rumah sakit. Jarak menuju tempat praktek dokter terdekat lebih dari atau sama dengan 14 km, dan untuk mencapai fasilitas tersebut dirasakan sulit atau sangat sulit.
- 2) **Kode 1 (satu)** merupakan desa yang tidak terdapat tempat praktek dokter, poliklinik/balai pengobatan, puskesmas, pustu, rumah sakit bersalin, maupun rumah sakit. Jarak menuju tempat praktek dokter terdekat lebih dari atau sama dengan 14 km, dan untuk mencapai fasilitas tersebut dirasakan mudah atau sangat mudah.
- 3) **Kode 2 (dua)** merupakan desa yang tidak terdapat tempat praktek dokter, poliklinik/balai pengobatan, puskesmas, pustu, rumah sakit bersalin, maupun rumah sakit. Jarak menuju tempat praktek dokter terdekat kurang dari 14 km, dan untuk mencapai fasilitas tersebut dirasakan sulit atau sangat sulit.
- 4) **Kode 3 (tiga)** merupakan desa yang tidak terdapat tempat praktek dokter, poliklinik/balai pengobatan, puskesmas, pustu, rumah sakit bersalin, maupun rumah sakit. Jarak menuju tempat praktek dokter

terdekat kurang dari 14 km, dan untuk mencapai fasilitas tersebut dirasakan mudah atau sangat mudah.

- 5) **Kode 4 (empat)** merupakan desa yang terdapat tempat praktek dokter dengan rasio ketersediaan tempat praktek dokter terhadap penduduk desa kurang dari 0,0003571413567 (4 fasilitas per 10.000 penduduk). Termasuk jika di desa tersebut terdapat poliklinik/balai pengobatan, tetapi tidak ada puskesmas, pustu, rumah sakit bersalin, maupun rumah sakit.
- 6) **Kode 5 (lima)** merupakan desa yang terdapat tempat praktek dokter dengan rasio ketersediaan tempat praktek dokter terhadap penduduk desa lebih dari atau sama dengan 0,0003571413567 (4 fasilitas per 10.000 penduduk). Termasuk jika di desa tersebut terdapat puskesmas, pustu, rumah sakit bersalin, atau rumah sakit, tanpa mempertimbangkan ketersediaan poliklinik/balai pengobatan.

Rekategorisasi:

- 1) Tidak tersedia tempat praktek dokter.
 - 2) Tersedia tempat praktek dokter.
11. Tempat praktek bidan adalah sarana kesehatan/bangunan yang digunakan untuk tempat praktek bidan yang biasanya memberikan pelayanan ibu hamil dan bayi.

Kategori dari Bappenas:

- 1) **Kode 0 (nol)** merupakan desa yang tidak terdapat tempat praktek bidan, tempat praktek dokter, poliklinik/balai pengobatan, puskesmas, pustu, rumah sakit bersalin, maupun rumah sakit. Jarak menuju tempat praktek bidan terdekat lebih dari atau sama dengan 16 km, dan untuk mencapai fasilitas tersebut dirasakan sulit atau sangat sulit.
- 2) **Kode 1 (satu)** merupakan desa yang tidak terdapat tempat praktek bidan, tempat praktek dokter, poliklinik/balai pengobatan, puskesmas, pustu, rumah sakit bersalin, maupun rumah sakit. Jarak menuju tempat praktek bidan terdekat lebih dari atau sama dengan 16 km, dan untuk mencapai fasilitas tersebut dirasakan mudah atau sangat mudah.

- 3) **Kode 2 (dua)** merupakan desa yang tidak terdapat tempat praktek bidan, tempat praktek dokter, poliklinik/balai pengobatan, puskesmas, pustu, rumah sakit bersalin, maupun rumah sakit. Jarak menuju tempat praktek bidan terdekat kurang dari 16 km, dan untuk mencapai fasilitas tersebut dirasakan sulit atau sangat sulit.
- 4) **Kode 3 (tiga)** merupakan desa yang tidak terdapat tempat praktek bidan, tempat praktek dokter, poliklinik/balai pengobatan, puskesmas, pustu, rumah sakit bersalin, maupun rumah sakit. Jarak menuju tempat praktek bidan terdekat kurang dari 16 km, dan untuk mencapai fasilitas tersebut dirasakan mudah atau sangat mudah.
- 5) **Kode 4 (empat)** merupakan desa yang terdapat tempat praktek bidan dengan rasio ketersediaan tempat praktek bidan terhadap penduduk desa kurang dari 0,00050732962398 (5 fasilitas per 10.000 penduduk). Termasuk jika di desa tersebut terdapat tempat praktek dokter, tetapi tidak tersedia poliklinik/balai pengobatan, puskesmas, pustu, rumah sakit bersalin, maupun rumah sakit.
- 6) **Kode 5 (lima)** merupakan desa yang terdapat tempat praktek bidan dengan rasio ketersediaan tempat praktek bidan terhadap penduduk desa lebih dari atau sama dengan 0,00050732962398 (5 fasilitas per 10.000 penduduk). Termasuk jika di desa tersebut terdapat poliklinik/balai pengobatan, puskesmas, pustu, rumah sakit bersalin, atau rumah sakit, tanpa mempertimbangkan ketersediaan tempat praktek dokter.

Rekategorisasi:

- 1) Tidak tersedia tempat praktek bidan.
 - 2) Tersedia tempat praktek bidan.
12. Pos Kesehatan Desa (Poskesdes) atau lebih sering dikenal sebagai PKD di beberapa wilayah merupakan sarana kesehatan/bangunan yang dibentuk di desa/kelurahan dalam rangka mendekatkan/menyediakan pelayanan kesehatan dasar bagi masyarakat desa/kelurahan.
- Pondok Bersalin Desa (Polindes) adalah bangunan yang dibangun dengan sumbangan dana pemerintah dan partisipasi masyarakat desa untuk tempat pertolongan persalinan dan pemondokan ibu bersalin, sekaligus tempat tinggal

bidan di desa. Di samping pertolongan persalinan juga dilakukan pelayanan Kesehatan Ibu dan Anak (KIA), Keluarga Berencana (KB), dan pelayanan kesehatan lain sesuai kebutuhan masyarakat dan kompetensi teknis bidan tersebut.

Kategori dari Bappenas:

- 1) **Kode 0 (nol)** merupakan desa yang tidak terdapat poskesdes, polindes, tempat praktek bidan, tempat praktek dokter, poliklinik/balai pengobatan, puskesmas, pustu, rumah sakit bersalin, maupun rumah sakit. Jarak menuju poskesdes atau polindes terdekat lebih dari atau sama dengan 8 km, dan untuk mencapai fasilitas tersebut dirasakan sulit atau sangat sulit.
- 2) **Kode 1 (satu)** merupakan desa yang tidak terdapat poskesdes, polindes, tempat praktek bidan, tempat praktek dokter, poliklinik/balai pengobatan, puskesmas, pustu, rumah sakit bersalin, maupun rumah sakit. Jarak menuju poskesdes atau polindes terdekat lebih dari atau sama dengan 8 km, dan untuk mencapai fasilitas tersebut dirasakan mudah atau sangat mudah.
- 3) **Kode 2 (dua)** merupakan desa yang tidak terdapat poskesdes, polindes, tempat praktek bidan, tempat praktek dokter, poliklinik/balai pengobatan, puskesmas, pustu, rumah sakit bersalin, maupun rumah sakit. Jarak menuju poskesdes atau polindes terdekat kurang dari 8 km, dan untuk mencapai fasilitas tersebut dirasakan sulit atau sangat sulit.
- 4) **Kode 3 (tiga)** merupakan desa yang tidak terdapat poskesdes, polindes, tempat praktek bidan, tempat praktek dokter, poliklinik/balai pengobatan, puskesmas, pustu, rumah sakit bersalin, maupun rumah sakit. Jarak menuju poskesdes atau polindes terdekat kurang dari 8 km, dan untuk mencapai fasilitas tersebut dirasakan mudah atau sangat mudah.
- 5) **Kode 4 (empat)** merupakan desa yang terdapat poskesdes atau polindes dengan rasio ketersediaan poskesdes atau polindes terhadap penduduk desa kurang dari 0,00070719038224 (7 fasilitas per 10.000 penduduk). Termasuk jika di desa tersebut terdapat tempat praktek bidan, tetapi tidak tersedia tempat praktek dokter, poliklinik/balai pengobatan, puskesmas, pustu, rumah sakit bersalin, maupun rumah sakit

- 6) **Kode 5 (lima)** merupakan desa yang terdapat poskesdes atau polindes dengan rasio ketersediaan poskesdes atau polindes terhadap penduduk desa lebih dari atau sama dengan 0,00070719038224 (7 fasilitas per 10.000 penduduk). Termasuk jika di desa tersebut terdapat tempat praktek dokter, poliklinik/balai pengobatan, puskesmas, pustu, rumah sakit bersalin, atau rumah sakit, tanpa mempertimbangkan ketersediaan tempat praktek bidan.

Rekategorisasi:

- 1) Tidak tersedia poskesdes atau polindes.
 - 2) Tersedia poskesdes atau polindes.
13. Apotek adalah suatu sarana kesehatan yang digunakan untuk pekerjaan kefarmasian, dan penyaluran/penjualan obat/bahan farmasi. Apotek melayani pembelian obat secara bebas atau dengan resep dokter. Apotek selalu ada tenaga apoteker selaku penanggungjawabnya.

Kategori dari Bappenas:

- 1) **Kode 0 (nol)** merupakan desa yang tidak terdapat apotek. Jarak menuju apotek terdekat lebih dari atau sama dengan 17 km, dan untuk mencapai fasilitas tersebut dirasakan sulit atau sangat sulit.
- 2) **Kode 1 (satu)** merupakan desa yang tidak terdapat apotek. Jarak menuju apotek terdekat lebih dari atau sama dengan 17 km, dan untuk mencapai fasilitas tersebut dirasakan mudah atau sangat mudah.
- 3) **Kode 2 (dua)** merupakan desa yang tidak terdapat apotek. Jarak menuju apotek terdekat kurang dari 17 km, dan untuk mencapai fasilitas tersebut dirasakan sulit atau sangat sulit.
- 4) **Kode 3 (tiga)** merupakan desa yang tidak terdapat apotek. Jarak menuju apotek terdekat kurang dari 17 km, dan untuk mencapai fasilitas tersebut dirasakan mudah atau sangat mudah.
- 5) **Kode 4 (empat)** merupakan desa yang terdapat apotek dengan rasio ketersediaan apotek terhadap penduduk desa kurang dari 0,00032420638537 (3 fasilitas per 10.000 penduduk).

- 6) **Kode 5 (lima)** merupakan desa yang terdapat apotek dengan rasio ketersediaan apotek terhadap penduduk desa lebih dari atau sama dengan 0,00032420638537 (3 fasilitas per 10.000 penduduk).

Rekategorisasi:

- 1) Tidak tersedia apotek.
- 2) Tersedia apotek.

14. Kelompok pertokoan adalah sejumlah toko yang terdiri dari minimal 10 toko dan mengelompok dalam satu lokasi. Dalam satu kelompok pertokoan, jumlah bangunan fisiknya bisa lebih dari satu.

Minimarket adalah sistem pelayanan mandiri, menjual berbagai jenis barang secara eceran, dan semua barang memiliki label harga, dengan luas bangunan kurang dari 400 m².

Toko/warung kelontong adalah bangunan yang berfungsi sebagai tempat usaha di bangunan tetap untuk menjual barang keperluan sehari-hari secara eceran, tidak mempunyai sistem pelayanan mandiri dikelola oleh satu penjual.

Kategori dari Bappenas:

- 1) **Kode 0 (nol)** merupakan desa yang tidak terdapat kelompok pertokoan, minimarket, maupun toko/warung kelontong.
- 2) **Kode 1 (satu)** merupakan desa yang tidak terdapat kelompok pertokoan maupun minimarket, tetapi tersedia toko/warung kelontong dengan rasio ketersediaan toko/warung kelontong terhadap penduduk desa kurang dari 0,01004450094485 (100 fasilitas per 10.000 penduduk).
- 3) **Kode 2 (dua)** merupakan desa yang tidak terdapat kelompok pertokoan maupun minimarket, tetapi tersedia toko/warung kelontong dengan rasio ketersediaan toko/warung kelontong terhadap penduduk desa lebih dari atau sama dengan 0,01004450094485 (100 fasilitas per 10.000 penduduk).
- 4) **Kode 3 (tiga)** merupakan desa yang tidak terdapat kelompok pertokoan tetapi tersedia minimarket dengan rasio ketersediaan minimarket terhadap penduduk desa kurang dari 0,00043681590218 (4 fasilitas per 10.000 penduduk), tanpa mempertimbangkan ketersediaan toko/warung kelontong.

- 5) **Kode 4 (empat)** merupakan desa yang tidak terdapat kelompok pertokoan tetapi tersedia minimarket dengan rasio ketersediaan minimarket terhadap penduduk desa lebih dari atau sama dengan 0,00043681590218 (4 fasilitas per 10.000 penduduk), tanpa mempertimbangkan ketersediaan toko/warung kelontong.
- 6) **Kode 5 (lima)** merupakan desa yang terdapat terdapat kelompok pertokoan tanpa mempertimbangkan ketersediaan minimarket maupun toko/warung kelontong.

Rekategorisasi:

- 1) Tidak tersedia pertokoan, minimarket, atau toko kelontong.
 - 2) Tersedia pertokoan, minimarket, atau toko kelontong.
15. Pasar adalah tempat pertemuan antara penjual dan pembeli barang dan jasa. Pasar bisa menggunakan bangunan yang bersifat permanen atau semi permanen ataupun tanpa bangunan. Barang yang diperjualbelikan di dalam pasar bisa terdiri dari banyak komoditas (campuran) ataupun secara khusus suatu komoditas tertentu. Contoh pasar yang secara khusus memang diperuntukkan untuk memperjualbelikan suatu komoditas tertentu adalah pasar ikan, pasar beras, dll.

Kategori dari Bappenas:

- 1) **Kode 0 (nol)** merupakan desa yang tidak terdapat pasar dengan bangunan permanen, pasar dengan bangunan semi permanen, maupun pasar tanpa bangunan.
- 2) **Kode 1 (satu)** merupakan desa yang tidak terdapat pasar dengan bangunan permanen maupun pasar dengan bangunan semi permanen, tetapi masih ada pasar tanpa bangunan.
- 3) **Kode 2 (dua)** merupakan desa yang tidak terdapat pasar dengan bangunan permanen, tetapi masih ada pasar dengan bangunan semi permanen dengan rasio ketersediaan pasar dengan bangunan semi permanen terhadap penduduk desa kurang dari 0,00053925247833 (5 fasilitas per 10.000 penduduk), tanpa mempertimbangkan ketersediaan pasar tanpa bangunan.
- 4) **Kode 3 (tiga)** merupakan desa yang tidak terdapat pasar dengan bangunan permanen, tetapi masih ada pasar dengan bangunan semi permanen dengan

rasio ketersediaan pasar dengan bangunan semi permanen terhadap penduduk desa lebih dari atau sama dengan 0,00053925247833 (5 fasilitas per 10.000 penduduk), tanpa mempertimbangkan ketersediaan pasar tanpa bangunan.

- 5) **Kode 4 (empat)** merupakan desa yang terdapat pasar dengan bangunan permanen dengan rasio ketersediaan pasar dengan bangunan permanen terhadap penduduk desa kurang dari 0,00039399640901 (4 fasilitas per penduduk), tanpa mempertimbangkan ketersediaan pasar dengan bangunan semi permanen maupun pasar tanpa bangunan.
- 6) **Kode 5 (lima)** merupakan desa yang terdapat pasar dengan bangunan permanen dengan rasio ketersediaan pasar dengan bangunan permanen terhadap penduduk desa lebih dari atau sama dengan 0,00039399640901 (4 fasilitas per 10.000 penduduk), tanpa mempertimbangkan ketersediaan pasar dengan bangunan semi permanen maupun pasar tanpa bangunan.

Rekategorisasi:

- 1) Tidak tersedia pasar.
 - 2) Tersedia pasar.
16. Restoran adalah suatu jenis usaha yang mempergunakan seluruh bangunan secara permanen untuk menyediakan jasa pangan yang pengolahan dan penyajiannya secara langsung di tempat sesuai dengan keinginan para pengguna jasa yang mempunyai ciri pembeli biasanya dikenakan pajak. Izin restoran dan kualifikasinya diberikan oleh Ditjen Pariwisata/Kanwil Parpostel setempat.

Rumah makan adalah jenis usaha yang menyediakan jasa pangan yang pengolahan makanannya bisa dilakukan diluar rumah makan, yang mempunyai ciri pembeli biasanya dikenakan pajak. Izin rumah makan diberikan oleh Diparda (pada kabupaten/kota). Di wilayah yang ada Dinas Pariwisata, biasanya pemberian izin ditangani oleh Direktorat Perekonomian/Bagian Perekonomian Pemda setempat.

Kategori dari Bappenas:

- 1) **Kode 0 (nol)** merupakan desa yang tidak terdapat restoran/rumah makan maupun warung/kedai makanan minuman.

- 2) **Kode 1 (satu)** merupakan desa yang tidak terdapat restoran/rumah makan, tetapi ada warung/kedai makanan minuman dengan rasio ketersediaan warung/kedai makanan minuman terhadap penduduk desa kurang dari 0,00408160850626 (40 fasilitas per 10.000 penduduk).
- 3) **Kode 2 (dua)** merupakan desa yang tidak terdapat restoran/rumah makan, tetapi ada warung/kedai makanan minuman dengan rasio ketersediaan warung/kedai makanan minuman terhadap penduduk desa lebih dari atau sama dengan 0,00408160850626 (40 fasilitas per 10.000 penduduk) dan kurang dari 0,00816321701253 (82 fasilitas per 10.000 penduduk).
- 4) **Kode 3 (tiga)** merupakan desa yang tidak terdapat restoran/rumah makan, tetapi ada warung/kedai makanan minuman dengan rasio ketersediaan warung/kedai makanan minuman terhadap penduduk desa lebih dari atau sama dengan 0,00816321701253 (82 fasilitas per 10.000 penduduk).
- 5) **Kode 4 (empat)** merupakan desa yang terdapat restoran/rumah makan dengan rasio ketersediaan restoran/rumah makan terhadap penduduk desa kurang dari 0,0007165878916 (72 fasilitas per 10.000 penduduk), tanpa mempertimbangkan ketersediaan warung/kedai makanan minuman.
- 6) **Kode 5 (lima)** merupakan desa yang terdapat restoran/rumah makan dengan rasio ketersediaan restoran/rumah makan terhadap penduduk desa lebih dari atau sama dengan 0,0007165878916 (72 fasilitas per 10.000 penduduk), tanpa mempertimbangkan ketersediaan warung/kedai makanan minuman.

Rekategorisasi:

- 1) Tidak tersedia restoran/rumah makan/warung makan.
 - 2) Tersedia restoran/rumah makan/warung makan.
17. Hotel adalah jenis akomodasi yang mempergunakan sebagian atau keseluruhan bangunan untuk jasa pelayanan penginapan, penyedia makanan dan minuman serta jasa lainnya (seperti restoran, binatu, dll) bagi masyarakat umum yang dikelola secara komersial dengan ijin usaha sebagai hotel.
- Penginapan (hostel/motel/losmen/wisma) adalah jenis akomodasi yang mempergunakan sebagian atau keseluruhan bangunan untuk jasa pelayanan

penginapan bagi umum, biasanya tanpa fasilitas pelayanan makan minum yang dikelola secara komersial dengan izin usaha bukan hotel.

Kategori dari Bappenas:

- 1) **Kode 0 (nol)** merupakan desa yang tidak terdapat hotel maupun penginapan.
- 2) **Kode 1 (satu)** merupakan desa yang tidak terdapat hotel tetapi ada penginapan dengan rasio ketersediaan penginapan terhadap penduduk desa kurang dari 0,00096503248013 (10 fasilitas per 10.000 penduduk).
- 3) **Kode 2 (dua)** merupakan desa yang tidak terdapat hotel tetapi ada penginapan dengan rasio ketersediaan penginapan terhadap penduduk desa lebih dari atau sama dengan 0,00096503248013 (10 fasilitas per 10.000 penduduk) dan kurang dari 0,00193006496027 (19 fasilitas per 10.000 penduduk).
- 4) **Kode 3 (tiga)** merupakan desa yang tidak terdapat hotel tetapi ada penginapan dengan rasio ketersediaan penginapan terhadap penduduk desa lebih dari atau sama dengan 0,00193006496027 (19 fasilitas per 10.000 penduduk).
- 5) **Kode 4 (empat)** merupakan desa yang terdapat hotel dengan rasio ketersediaan hotel terhadap penduduk desa kurang dari 0,00044619619601 (4 fasilitas per 10.000 penduduk), tanpa mempertimbangkan ketersediaan penginapan.
- 6) **Kode 5 (lima)** merupakan desa yang terdapat hotel dengan rasio ketersediaan hotel terhadap penduduk desa lebih dari atau sama dengan 0,00044619619601 (4 fasilitas per 10.000 penduduk), tanpa mempertimbangkan ketersediaan penginapan.

Rekategorisasi:

- 1) Tidak tersedia akomodasi hotel/penginapan.
 - 2) Tersedia akomodasi hotel/penginapan.
18. Ketersediaan bank mengacu pada keberadaan bank umum, baik bank umum pemerintah maupun swasta dan Bank Perkreditan Rakyat yang masih aktif/beroperasi di wilayah desa.

Bank Umum adalah bank yang dapat memberikan jasa dalam lalu lintas pembayaran. Usaha dari bank umum adalah menghimpun dana masyarakat dalam bentuk giro, deposito berjangka, sertifikat deposito dan tabungan serta menyalurkan kredit

Bank Perkreditan Rakyat (BPR) adalah bank yang menerima simpanan dalam bentuk deposito berjangka, tabungan atau bentuk lain yang disamakan dengan itu, menyalurkan dana dalam bentuk kredit kepada masyarakat yang membutuhkan.

Kategori dari Bappenas:

- 1) **Kode 0 (nol)** merupakan desa yang tidak terdapat bank.
- 2) **Kode 1 (satu)** merupakan desa yang terdapat bank dengan rasio ketersediaan bank terhadap penduduk desa kurang dari 0.00022567912721 (2 fasilitas per 10.000 penduduk).
- 3) **Kode 2 (dua)** merupakan desa yang terdapat bank dengan rasio ketersediaan bank terhadap penduduk desa lebih dari atau sama dengan 0.00022567912721 (2 fasilitas per 10.000 penduduk) dan kurang dari 0,00045135825442 (5 fasilitas per 10.000 penduduk).
- 4) **Kode 3 (tiga)** merupakan desa yang terdapat bank dengan rasio ketersediaan bank terhadap penduduk desa lebih dari atau sama dengan 0,00045135825442 (5 fasilitas per 10.000 penduduk) dan kurang dari 0,00067703738163 (7 fasilitas per 10.000 penduduk).
- 5) **Kode 4 (empat)** merupakan desa yang terdapat bank dengan rasio ketersediaan bank terhadap penduduk desa lebih dari atau sama dengan 0,00067703738163 (7 fasilitas per 10.000 penduduk) dan kurang dari 0,00090271650884 (9 fasilitas per 10.000 penduduk).
- 6) **Kode 5 (lima)** merupakan desa yang terdapat bank dengan rasio ketersediaan bank terhadap penduduk desa lebih dari atau sama dengan 0,00090271650884 (9 fasilitas per 10.000 penduduk).

Rekategorisasi:

- 1) Tidak tersedia bank.
- 2) Tersedia bank.

19. Elektrifikasi mengacu pada persentase keluarga pengguna listrik.

Keluarga pengguna listrik Perusahaan Listrik Negara (PLN) adalah keluarga pengguna/pelanggan listrik yang disalurkan oleh PLN.

Keluarga pengguna listrik non-PLN adalah keluarga pengguna/pelanggan listrik selain dari PLN, misalnya diesel/generator, listrik diusahakan oleh pemerintah daerah, swasta, dan listrik swadaya masyarakat.

Kategori dari Bappenas:

- 1) **Kode 0 (nol)** merupakan desa yang persentase keluarga pengguna listrik kurang dari atau sama dengan 17 persen.
- 2) **Kode 1 (satu)** merupakan desa yang persentase keluarga pengguna listrik lebih dari 17 persen dan kurang dari atau sama dengan 33 persen.
- 3) **Kode 2 (dua)** merupakan desa yang persentase keluarga pengguna listrik lebih dari 33 persen dan kurang dari atau sama dengan 50 persen.
- 4) **Kode 3 (tiga)** merupakan desa yang persentase keluarga pengguna listrik lebih dari 50 persen dan kurang dari atau sama dengan 67 persen.
- 5) **Kode 4 (empat)** merupakan desa yang persentase keluarga pengguna listrik lebih dari 67 persen dan kurang dari atau sama dengan 83 persen.
- 6) **Kode 5 (lima)** merupakan desa yang persentase keluarga pengguna listrik lebih dari 83 persen.

Rekategorisasi:

- 1) Persentase keluarga pengguna listrik kurang dari atau sama dengan 80 persen.
- 2) Persentase keluarga pengguna listrik lebih dari 80 persen.

20. Kondisi penerangan di jalan utama dibedakan menjadi tidak ada penerangan, penerangan non listrik, dan penerangan listrik.

Kategori dari Bappenas:

- 1) **Kode 0 (nol)** merupakan desa yang tidak ada penerangan di jalan utama desa.
- 2) **Kode 2 (dua)** merupakan desa yang ada penerangan di jalan utama desa dengan jenis penerangannya berupa non listrik.

- 3) **Kode 3 (tiga)** merupakan desa yang ada penerangan di jalan utama desa dengan jenis penerangannya berupa listrik yang diusahakan oleh non pemerintah.
- 4) **Kode 5 (lima)** merupakan desa yang ada penerangan di jalan utama desa dengan jenis penerangannya berupa listrik yang diusahakan oleh pemerintah.

Rekategorisasi:

- 1) Tidak ada penerangan di jalan utama desa.
- 2) Ada penerangan di jalan utama desa.

21. Bahan bakar untuk memasak mengacu pada jenis bahan bakar untuk memasak yang digunakan oleh mayoritas keluarga di desa.

Kategori dari Bappenas:

- 1) **Kode 0 (nol)** merupakan desa yang sebagian besar keluarganya menggunakan batubara, arang, dan lain-lain untuk memasak.
- 2) **Kode 1 (satu)** merupakan desa yang sebagian besar keluarganya menggunakan kayu bakar untuk memasak.
- 3) **Kode 2 (dua)** merupakan desa yang sebagian besar keluarganya menggunakan minyak tanah untuk memasak, tetapi tidak ada agen/penjual minyak tanah.
- 4) **Kode 3 (tiga)** merupakan desa yang sebagian besar keluarganya menggunakan minyak tanah untuk memasak, dan ada agen/penjual minyak tanah.
- 5) **Kode 4 (empat)** merupakan desa yang sebagian besar keluarganya menggunakan LPG atau gas kota untuk memasak, tetapi tidak ada pangkalan/agen/penjual LPG.
- 6) **Kode 5 (lima)** merupakan desa yang sebagian besar keluarganya menggunakan LPG atau gas kota, dan ada pangkalan/agen/penjual LPG.

Rekategorisasi:

- 1) Penggunaan mayoritas adalah minyak tanah, kayu bakar, batu bara, dll.
- 2) Penggunaan mayoritas adalah LPG atau gas kota.

22. Sumber air minum mengacu kepada jenis sumber air minum yang digunakan oleh mayoritas keluarga di desa.

Kategori dari Bappenas:

- 1) **Kode 0 (nol)** merupakan desa yang sumber air untuk minum sebagian besar keluarga berasal dari air hujan atau lainnya.
- 2) **Kode 1 (satu)** merupakan desa yang sumber air untuk minum sebagian besar keluarga berasal dari sungai/danau/kolam.
- 3) **Kode 2 (dua)** merupakan desa yang sumber air untuk minum sebagian besar keluarga berasal dari mata air.
- 4) **Kode 3 (tiga)** merupakan desa yang sumber air untuk minum sebagian besar keluarga berasal dari sumur.
- 5) **Kode 4 (empat)** merupakan desa yang sumber air untuk minum sebagian besar keluarga berasal dari ledeng tanpa meteran, sumur bor, atau pompa.
- 6) **Kode 5 (lima)** merupakan desa yang sumber air untuk minum sebagian besar keluarga berasal dari air kemasan, atau ledeng dengan meteran.

Rekategorisasi:

- 1) Penggunaan mayoritas adalah air hujan, sungai/danau/kolam, atau mata air.
 - 2) Penggunaan mayoritas adalah sumur, ledeng, atau air kemasan.
23. Sumber air untuk mandi/cuci mengacu kepada jenis sumber air yang digunakan oleh mayoritas keluarga di desa untuk mandi/cuci.

Kategori dari Bappenas:

- 1) **Kode 0 (nol)** merupakan desa yang sumber air untuk mandi/cuci sebagian besar keluarga berasal dari air hujan atau lainnya.
- 2) **Kode 1 (satu)** merupakan desa yang sumber air untuk mandi/cuci sebagian besar keluarga berasal dari sungai/danau/kolam.
- 3) **Kode 2 (dua)** merupakan desa yang sumber air untuk mandi/cuci sebagian besar keluarga berasal dari mata air.
- 4) **Kode 3 (tiga)** merupakan desa yang sumber air untuk mandi/cuci sebagian besar keluarga berasal dari sumur.
- 5) **Kode 4 (empat)** merupakan desa yang sumber air untuk mandi/cuci sebagian besar keluarga berasal dari ledeng tanpa meteran, sumur bor, atau pompa.

- 6) **Kode 5 (lima)** merupakan desa yang sumber air untuk mandi/cuci sebagian besar keluarga berasal dari ledeng dengan meteran.

Rekategorisasi:

- 1) Penggunaan mayoritas adalah air hujan, sungai/danau/kolam, atau mata air.
- 2) Penggunaan mayoritas adalah sumur atau ledeng.

24. Fasilitas buang air besar mengacu pada jenis fasilitas yang digunakan oleh mayoritas keluarga di desa.

- a. Jamban adalah tempat buang air besar yang tertutup, baik menggunakan tangki septik maupun tidak.
- b. Bukan jamban termasuk tempat pembuangan air besar yang penampungan akhirnya kolam/sawah, lubang tanah/tanah lapang/kebun, sungai/danau/laut, dan sebagainya.

Kategori dari Bappenas:

- 1) **Kode 0 (nol)** merupakan desa yang fasilitas buang air besar sebagian besar keluarga adalah bukan jamban.
- 2) **Kode 2 (dua)** merupakan desa yang fasilitas buang air besar sebagian besar keluarga adalah jamban umum.
- 3) **Kode 3 (tiga)** merupakan desa yang fasilitas buang air besar sebagian besar keluarga adalah jamban bersama.
- 4) **Kode 5 (lima)** merupakan desa yang fasilitas buang air besar sebagian besar keluarga adalah jamban sendiri.

Rekategorisasi:

- 1) Bukan jamban.
- 2) Jamban.

25. Kualitas fasilitas komunikasi seluler ditandai dengan adanya sinyal telepon seluler. Sinyal telepon seluler adalah besaran elektromagnetik yang berubah dalam ruang dan waktu dengan membawa informasi yang memberikan konfirmasi bahwa layanan telepon seluler sudah tersedia. Rincian ini terdiri dari tidak ada sinyal, sinyal lemah, dan sinyal kuat.

Kategori dari Bappenas:

- 1) **Kode 0 (nol)** merupakan desa yang tidak ada *Base Transceiver Station* (BTS) dan tidak ada sinyal telepon seluler/*handphone*.
- 2) **Kode 1 (satu)** merupakan desa yang ada *Base Transceiver Station* (BTS), tetapi tidak ada sinyal telepon seluler/*handphone*.
- 3) **Kode 2 (dua)** merupakan desa yang tidak ada *Base Transceiver Station* (BTS), tetapi ada sinyal telepon seluler/*handphone* yang lemah.
- 4) **Kode 3 (tiga)** merupakan desa yang ada *Base Transceiver Station* (BTS), tetapi ada sinyal telepon seluler/*handphone* yang lemah.
- 5) **Kode 4 (empat)** merupakan desa yang tidak ada *Base Transceiver Station* (BTS), tetapi ada sinyal telepon seluler/*handphone* yang kuat.
- 6) **Kode 5 (lima)** merupakan desa yang ada *Base Transceiver Station* (BTS), tetapi ada sinyal telepon seluler/*handphone* yang kuat.

Rekategorisasi:

- 1) Tidak ada sinyal telepon seluler.
 - 2) Ada sinyal telepon seluler.
26. Ketersediaan fasilitas internet ditandai dengan keberadaan fasilitas internet di kantor kepala desa atau warnet di desa.

Kantor pos adalah pemberi pelayanan komunikasi tertulis dan/atau surat elektronik, layanan paket, layanan logistik, layanan transaksi keuangan, dan layanan keagenan pos untuk kepentingan umum.

Jasa pengiriman paket/dokumen swasta adalah pelayanan pengiriman paket maupun dokumen yang dikelola oleh pihak swasta, misalnya Tiki, JNE, ESL, dll.

Kategori dari Bappenas:

- 1) **Kode 0 (nol)** merupakan desa yang tidak ada kelima jenis fasilitas yaitu: internet di kantor kepala desa, warnet, kantor pos/pos pembantu/rumah pos, pos keliling, serta jasa ekspedisi.
- 2) **Kode 1 (satu)** merupakan desa yang ada satu jenis dari kelima fasilitas yaitu: internet di kantor kepala desa, warnet, kantor pos/pos pembantu/rumah pos, pos keliling, serta jasa ekspedisi.

- 3) **Kode 2 (dua)** merupakan desa yang ada dua jenis diantara kelima fasilitas yaitu: internet di kantor kepala desa, warnet, kantor pos/pos pembantu/rumah pos, pos keliling, serta jasa ekspedisi.
- 4) **Kode 3 (tiga)** merupakan desa yang ada tiga jenis diantara kelima fasilitas yaitu: internet di kantor kepala desa, warnet, kantor pos/pos pembantu/rumah pos, pos keliling, serta jasa ekspedisi.
- 5) **Kode 4 (empat)** merupakan desa yang ada empat jenis diantara kelima fasilitas yaitu: internet di kantor kepala desa, warnet, kantor pos/pos pembantu/rumah pos, pos keliling, serta jasa ekspedisi.
- 6) **Kode 5 (lima)** merupakan desa yang ada kelima jenis fasilitas yaitu: internet di kantor kepala desa, warnet, kantor pos/pos pembantu/rumah pos, pos keliling, serta jasa ekspedisi.

Rekategorisasi:

- 1) Tidak tersedia fasilitas internet atau pengiriman pos/barang.
- 2) Tersedia fasilitas internet dan pengiriman pos/barang.

27. Lalu lintas dan kualitas jalan memuat keterangan sebagai berikut:

Kategori dari Bappenas:

- 1) **Kode 1 (satu)** merupakan desa yang lalu lintas dari dan ke desa hanya melalui air.
- 2) **Kode 2 (dua)** merupakan desa yang lalu lintas dari dan ke desa melalui darat, serta jenis permukaan jalan terluasnya selain aspal/beton, diperkeras, maupun tanah.
- 3) **Kode 3 (tiga)** merupakan desa yang lalu lintas dari dan ke desa melalui darat, serta jenis permukaan jalan terluasnya adalah tanah.
- 4) **Kode 4 (empat)** merupakan desa yang lalu lintas dari dan ke desa melalui darat, serta jenis permukaan jalan terluasnya adalah diperkeras (batu, kerikil, dan lain-lain).
- 5) **Kode 5 (lima)** merupakan desa yang lalu lintas dari dan ke desa melalui darat, serta jenis permukaan jalan terluasnya adalah aspal/beton.

Rekategorisasi:

- 1) Lalu lintas melalui air atau darat dengan jenis permukaan jalan terluasnya adalah selain aspal/beton.

- 2) Lalu lintas melalui darat dengan jenis permukaan jalan terluasnya adalah aspal/beton.

28. Aksesibilitas jalan memuat keterangan sebagai berikut:

Kategori dari Bappenas:

- 1) **Kode 1 (satu)** merupakan desa yang lalu lintas dari dan ke desa hanya melalui air.
- 2) **Kode 2 (dua)** merupakan desa yang lalu lintas dari dan ke desa melalui darat, atau darat dan air, tetapi jalannya tidak dapat dilalui kendaraan bermotor roda 4 atau lebih sepanjang tahun.
- 3) **Kode 3 (tiga)** merupakan desa yang lalu lintas dari dan ke desa melalui darat, atau darat dan air, serta jalannya dapat dilalui kendaraan bermotor roda 4 atau lebih sepanjang tahun kecuali sepanjang musim hujan.
- 4) **Kode 4 (empat)** merupakan desa yang lalu lintas dari dan ke desa melalui darat, atau darat dan air, serta jalannya dapat dilalui kendaraan bermotor roda 4 atau lebih sepanjang tahun kecuali saat tertentu (ketika turun hujan, pasang, dan lain-lain).
- 5) **Kode 5 (lima)** merupakan desa yang lalu lintas dari dan ke desa melalui darat, atau darat dan air, dan jalannya dapat dilalui kendaraan bermotor roda 4 atau lebih sepanjang tahun.

Rekategorisasi:

- 1) Lalu lintas melalui air atau darat dimana jalannya tidak dapat dilalui kendaraan bermotor roda 4 atau lebih sepanjang tahun.
- 2) Lalu lintas dari darat, atau darat dan air dimana jalannya dapat dilalui kendaraan bermotor roda 4 atau lebih sepanjang tahun.

29. Angkutan adalah suatu kegiatan usaha menyediakan jasa angkutan penumpang dan atau barang/ternak dari suatu tempat ke tempat lain dengan menggunakan alat angkutan bermotor maupun tidak bermotor, baik melalui darat, air maupun udara.

Angkutan umum ditandai dengan adanya trayek angkutan. Trayek angkutan adalah lintasan/rute/jalur angkutan umum untuk pelayanan jasa angkutan orang, barang dan atau orang dan barang yang mempunyai asal, tujuan dan lintasan perjalanan yang tetap tidak termasuk hanya barang.

Kategori dari Bappenas:

- 1) **Kode 0 (nol)** merupakan desa yang lalu lintas dari dan ke desa hanya melalui air.
- 2) **Kode 1 (satu)** merupakan desa yang jalannya tidak dilintasi oleh angkutan umum.
- 3) **Kode 2 (dua)** merupakan desa yang jalannya dilintasi oleh angkutan umum tanpa trayek tetap yang operasional angkutannya tidak setiap hari.
- 4) **Kode 3 (tiga)** merupakan desa yang jalannya dilintasi oleh angkutan umum tanpa trayek tetap yang operasional angkutannya setiap hari.
- 5) **Kode 4 (empat)** merupakan desa yang jalannya dilintasi oleh angkutan umum dengan trayek tetap yang operasional angkutannya tidak setiap hari.
- 6) **Kode 5 (lima)** merupakan desa yang jalannya dilintasi oleh angkutan umum dengan trayek tetap yang operasional angkutannya setiap hari.

Rekategorisasi:

- 1) Tidak ada angkutan umum.
 - 2) Ada angkutan umum.
30. Operasional angkutan umum mengacu pada ketersediaan angkutan umum setiap hari atau tidak setiap hari.

Kategori dari Bappenas:

- 1) **Kode 0 (nol)** merupakan desa yang lalu lintas dari dan ke desa hanya melalui air.
- 2) **Kode 1 (satu)** merupakan desa yang jalannya tidak dilewati oleh angkutan umum.
- 3) **Kode 2 (dua)** merupakan desa yang jalannya dilewati oleh angkutan umum dengan trayek yang tidak tetap yang jam operasionalnya hanya siang hari.
- 4) **Kode 3 (tiga)** merupakan desa yang jalannya dilewati oleh angkutan umum dengan trayek yang tidak tetap yang jam operasionalnya siang dan malam hari.
- 5) **Kode 4 (empat)** merupakan desa yang jalannya dilewati oleh angkutan umum dengan trayek tetap yang jam operasionalnya hanya siang hari.

- 6) **Kode 5 (lima)** merupakan desa yang jalannya dilewati oleh angkutan umum dengan trayek tetap yang jam operasionalnya siang dan malam hari.

Rekategorisasi:

- 1) Tidak tersedia setiap hari.
- 2) Tersedia setiap hari.

31. Waktu tempuh yang dicatat adalah rata-rata waktu tempuh dengan kendaraan yang biasanya digunakan oleh warga ke kantor Camat.

Kategori dari Bappenas:

- 1) **Kode 0 (nol)** merupakan desa yang waktu tempuh per kilometer ke kantor camatnya lebih besar atau sama dengan 0,76948044373697 jam/km (46 menit/km).
- 2) **Kode 1 (satu)** merupakan desa yang waktu tempuh per kilometer ke kantor camatnya lebih dari atau sama dengan 0,61558435498957 jam/km (36 menit/km) dan kurang dari 0,76948044373697 jam/km (46 menit/km).
- 3) **Kode 2 (dua)** merupakan desa yang waktu tempuh per kilometer ke kantor camatnya lebih dari atau sama dengan 0,46168826624218 jam/km (27 menit/km) dan kurang dari 0,61558435498957 jam/km (36 menit/km).
- 4) **Kode 3 (tiga)** merupakan desa yang waktu tempuh per kilometer ke kantor camatnya lebih dari atau sama dengan 0,30779217749479 jam/km (18 menit/km) dan kurang dari 0,46168826624218 jam/km (27 menit/km).
- 5) **Kode 4 (empat)** merupakan desa yang waktu tempuh per kilometer ke kantor camatnya lebih dari atau sama dengan 0,15389608874739 jam/km (9 menit/km) dan kurang dari 0,30779217749479 jam/km (18 menit/km).
- 6) **Kode 5 (lima)** merupakan desa yang waktu tempuh per kilometer ke kantor camatnya kurang dari 0,15389608874739 jam/km (9 menit/km).

Rekategorisasi:

- 1) Waktu tempuh lebih dari 9 menit/km.
- 2) Waktu tempuh kurang dari atau sama dengan 9 menit/km.

32. Biaya transportasi adalah rata-rata biaya yang dikeluarkan untuk sekali jalan. Bila rute yang digunakan pulang dan pergi berbeda maka yang digunakan adalah biaya rata-rata. Biaya per kilometer ke kantor camat adalah biaya transportasi per kilometer warga untuk menuju kantor Camat.

Kategori dari Bappenas:

- 1) **Kode 0 (nol)** merupakan desa yang biaya per kilometer ke kantor camatnya lebih dari atau sama dengan Rp. 35.000,00.
- 2) **Kode 1 (satu)** merupakan desa yang biaya per kilometer ke kantor camatnya lebih dari atau sama dengan Rp. 28.000,00 dan kurang dari Rp. 35.000,00.
- 3) **Kode 2 (dua)** merupakan desa yang biaya per kilometer ke kantor camatnya lebih dari atau sama dengan Rp. 21.000,00 dan kurang dari Rp. 28.000,00.
- 4) **Kode 3 (tiga)** merupakan desa yang biaya per kilometer ke kantor camatnya lebih dari atau sama dengan Rp. 14.000,00 dan kurang dari Rp. 21.000,00.
- 5) **Kode 4 (empat)** merupakan desa yang biaya per kilometer ke kantor camatnya lebih dari atau sama dengan Rp. 7.000,00 dan kurang dari Rp. 14.000,00.
- 6) **Kode 5 (lima)** merupakan desa yang biaya per kilometer ke kantor camatnya kurang dari Rp. 7.000,00.

Rekategorisasi:

- 1) Biaya lebih dari Rp. 7.000,00/km.
 - 2) Biaya kurang dari atau sama dengan Rp. 7.000,00/km.
33. Waktu tempuh yang dicatat adalah rata-rata waktu tempuh dengan kendaraan yang biasanya digunakan oleh warga ke kantor Bupati/Walikota.

Kategori dari Bappenas:

- 1) **Kode 0 (nol)** merupakan desa yang waktu tempuh per kilometer ke kantor bupati/walikotanya lebih dari atau sama dengan 0,19149877356422 (11 menit/km).
- 2) **Kode 1 (satu)** merupakan desa yang waktu tempuh per kilometer ke kantor bupati/walikotanya lebih dari atau sama dengan 0,15319901885137 (9 menit/km) dan kurang dari 0,19149877356422 (11 menit/km).
- 3) **Kode 2 (dua)** merupakan desa yang waktu tempuh per kilometer ke kantor bupati/walikotanya lebih dari atau sama dengan 0,11489926413853 (7 menit/km) dan kurang dari 0,15319901885137 (9 menit/km).

- 4) **Kode 3 (tiga)** merupakan desa yang waktu tempuh per kilometer ke kantor bupati/walikotanya lebih dari atau sama dengan 0,07659950942569 (5 menit/km) dan kurang dari 0,11489926413853 (7 menit/km).
- 5) **Kode 4 (empat)** merupakan desa yang waktu tempuh per kilometer ke kantor bupati/walikotanya lebih dari atau sama dengan 0,03829975471284 (2 menit/km) dan kurang dari 0,07659950942569 (5 menit/km).
- 6) **Kode 5 (lima)** merupakan desa yang waktu tempuh per kilometer ke kantor bupati/walikotanya kurang dari 0,03829975471284 (2 menit/km).

Rekategorisasi:

- 1) Waktu tempuh lebih dari 5 menit/km.
 - 2) Waktu tempuh kurang dari atau sama dengan 5 menit/km.
34. Biaya per kilometer ke kantor Bupati/Walikota adalah biaya transportasi per kilometer warga untuk menuju kantor Bupati/Walikota.

Kategori dari Bappenas:

- 1) **Kode 0 (nol)** merupakan desa yang biaya per kilometer ke kantor bupati/walikotanya lebih dari atau sama dengan Rp. 16.000,00.
- 2) **Kode 1 (satu)** merupakan desa yang biaya per kilometer ke kantor bupati/walikotanya lebih dari atau sama dengan Rp. 13.000,00 dan kurang dari Rp. 16.000,00.
- 3) **Kode 2 (dua)** merupakan desa yang biaya per kilometer ke kantor bupati/walikotanya lebih dari atau sama dengan Rp. 10.000,00 dan kurang dari Rp. 13.000,00.
- 4) **Kode 3 (tiga)** merupakan desa yang biaya per kilometer ke kantor bupati/walikotanya lebih dari atau sama dengan Rp. 6.000,00 dan kurang dari Rp. 10.000,00.
- 5) **Kode 4 (empat)** merupakan desa yang biaya per kilometer ke kantor bupati/walikotanya lebih dari atau sama dengan Rp. 3.000,00 dan kurang dari Rp. 6.000,00.
- 6) **Kode 5 (lima)** merupakan desa yang biaya per kilometer ke kantor bupati/walikotanya kurang dari Rp. 3.000,00.

Rekategorisasi:

- 1) Biaya lebih dari Rp. 3.000,00/km.
- 2) Biaya kurang dari atau sama dengan Rp. 3.000,00/km.

35. Penanganan kejadian luar biasa (KLB)

Kategori dari Bappenas:

- 1) **Kode 0 (nol)** merupakan desa ada kejadian luar biasa (KLB) atau wabah penyakit. Rasio penderita yang meninggal terhadap penderita lebih dari atau sama dengan 0,31400410439919 (32 orang meninggal per 100 penderita).
- 2) **Kode 1 (satu)** merupakan desa ada kejadian luar biasa (KLB) atau wabah penyakit. Rasio penderita yang meninggal terhadap penderita lebih dari atau sama dengan 0,15700205219959 (16 orang meninggal per 100 penderita) dan kurang dari 0,31400410439919 (32 orang meninggal per 100 penderita).
- 3) **Kode 2 (dua)** merupakan desa ada kejadian luar biasa (KLB) atau wabah penyakit. Rasio penderita yang meninggal terhadap penderita kurang dari 0,15700205219959 (16 orang meninggal per 100 penderita).
- 4) **Kode 3 (tiga)** merupakan desa yang ada kejadian luar biasa (KLB) atau wabah penyakit, tetapi tidak ada penderita yang meninggal
- 5) **Kode 5 (lima)** merupakan desa yang tidak ada kejadian luar biasa (KLB) atau wabah penyakit.

Rekategorisasi:

- 1) Ada KLB.
- 2) Tidak ada KLB.

36. Penanganan gizi buruk

Kategori dari Bappenas:

- 1) **Kode 0 (nol)** merupakan desa terdapat penderita gizi buruk dengan rasio penderita gizi buruk terhadap penduduk desa lebih dari 0,00510244452603 (5 penderita per 1000 penduduk).
- 2) **Kode 1 (satu)** merupakan desa terdapat penderita gizi buruk dengan rasio penderita gizi buruk terhadap penduduk desa lebih dari atau sama dengan

0,00340162968402 (3 penderita per 1000 penduduk) dan kurang dari 0,00510244452603 (5 penderita per 1000 penduduk).

- 3) **Kode 2 (dua)** merupakan desa terdapat penderita gizi buruk dengan rasio penderita gizi buruk terhadap penduduk desa lebih dari atau sama dengan 0,00170081484201 (2 penderita per 1000 penduduk) dan kurang dari 0,00340162968402 (3 penderita per 1000 penduduk).
- 4) **Kode 3 (tiga)** merupakan desa terdapat penderita gizi buruk dengan rasio penderita gizi buruk terhadap penduduk desa kurang dari 0,00170081484201 (2 penderita per 1000 penduduk).
- 5) **Kode 5 (lima)** merupakan desa yang tidak terdapat penderita gizi buruk.

Rekategorisasi:

- 1) Terdapat penderita gizi buruk.
- 2) Tidak terdapat penderita gizi buruk.

37. Fasilitas atau lapangan olahraga adalah tempat lapang untuk kegiatan olahraga yang ada di desa/kelurahan sesuai dengan persyaratan olahraga yang bersangkutan. Keberadaan lapangan olahraga yang dimaksudkan bukan hanya yang dimiliki oleh desa/kelurahan, termasuk lapangan yang dimiliki swasta atau pribadi yang difungsikan secara komersial dan masyarakat umum dapat mengaksesnya.

Kategori dari Bappenas:

- 1) **Kode 0 (nol)** merupakan desa yang tidak tersedia fasilitas/lapangan olahraga apapun (sepak bola, bola voli, bulu tangkis, bola basket, tenis lapangan, tenis meja, futsal, renang, bela diri, bilyard).
- 2) **Kode 1 (satu)** merupakan desa yang tersedia 1 atau 2 jenis fasilitas/lapangan olah raga (sepak bola, bola voli, bulu tangkis, bola basket, tenis lapangan, tenis meja, futsal, renang, bela diri, bilyard).
- 3) **Kode 2 (dua)** merupakan desa yang tersedia 3 atau 4 jenis fasilitas/lapangan olah raga (sepak bola, bola voli, bulu tangkis, bola basket, tenis lapangan, tenis meja, futsal, renang, bela diri, bilyard).
- 4) **Kode 3 (tiga)** merupakan desa yang tersedia 5 atau 6 jenis fasilitas/lapangan olah raga (sepak bola, bola voli, bulu tangkis, bola basket, tenis lapangan, tenis meja, futsal, renang, bela diri, bilyard).

- 5) **Kode 4 (empat)** merupakan desa yang tersedia 7 atau 8 jenis fasilitas/lapangan olah raga (sepak bola, bola voli, bulu tangkis, bola basket, tenis lapangan, tenis meja, futsal, renang, bela diri, bilyard).
- 6) **Kode 5 (lima)** merupakan desa yang tersedia 9 atau 10 jenis fasilitas/lapangan olah raga (sepak bola, bola voli, bulu tangkis, bola basket, tenis lapangan, tenis meja, futsal, renang, bela diri, bilyard).

Rekategorisasi:

- 1) Tidak tersedia fasilitas atau lapangan olahraga.
- 2) Tersedia fasilitas atau lapangan olahraga.

38. Kelompok kegiatan olah raga mengacu pada kelompok kegiatan yang dibentuk oleh warga desa dan anggotanya adalah warga desa/kelurahan setempat maupun warga di luar desa/kelurahan, tanpa memperhatikan apakah kegiatan olahraga tersebut dilakukan di desa/kelurahan maupun di tempat lain.

Kategori dari Bappenas:

- 1) **Kode 0 (nol)** merupakan desa yang tidak ada kelompok kegiatan olah raga apapun (sepak bola, bola voli, bulu tangkis, bola basket, tenis lapangan, tenis meja, futsal, renang, bela diri, bilyard).
- 2) **Kode 1 (satu)** merupakan desa yang ada 1 atau 2 jenis kelompok kegiatan olah raga (sepak bola, bola voli, bulu tangkis, bola basket, tenis lapangan, tenis meja, futsal, renang, bela diri, bilyard).
- 3) **Kode 2 (dua)** merupakan desa yang ada 3 atau 4 jenis kelompok kegiatan olah raga (sepak bola, bola voli, bulu tangkis, bola basket, tenis lapangan, tenis meja, futsal, renang, bela diri, bilyard).
- 4) **Kode 3 (tiga)** merupakan desa yang ada 5 atau 6 jenis kelompok kegiatan olah raga (sepak bola, bola voli, bulu tangkis, bola basket, tenis lapangan, tenis meja, futsal, renang, bela diri, bilyard).
- 5) **Kode 4 (empat)** merupakan desa yang ada 7 atau 8 jenis kelompok kegiatan olah raga (sepak bola, bola voli, bulu tangkis, bola basket, tenis lapangan, tenis meja, futsal, renang, bela diri, bilyard).
- 6) **Kode 5 (lima)** merupakan desa yang ada 9 atau 10 jenis kelompok kegiatan olah raga (sepak bola, bola voli, bulu tangkis, bola basket, tenis lapangan, tenis meja, futsal, renang, bela diri, bilyard).

Rekategorisasi:

- 1) Tidak ada kelompok kegiatan olah raga.
- 2) Ada kelompok kegiatan olah raga.

39. Kelengkapan pemerintahan desa mengacu pada keberadaan BPD dan kantor kepala desa.

Kategori dari Bappenas:

- 1) **Kode 0 (nol)** merupakan desa yang tidak ada BPD dan tidak ada kantor kepala desa. Tanpa mempertimbangkan ketersediaan batas wilayah desa dalam bentuk peta yang ditetapkan oleh bupati/walikota.
- 2) **Kode 1 (satu)** merupakan desa yang ada BPD tetapi tidak ada kantor kepala desa. Tanpa mempertimbangkan ketersediaan batas wilayah desa dalam bentuk peta yang ditetapkan oleh bupati/walikota.
- 3) **Kode 2 (dua)** merupakan desa yang tidak ada BPD tetapi ada kantor kepala desa di luar wilayah desa. Tanpa mempertimbangkan ketersediaan batas wilayah desa dalam bentuk peta yang ditetapkan oleh bupati/walikota.
- 4) **Kode 3 (tiga)** merupakan desa yang ada BPD dan ada kantor kepala desa di luar wilayah desa. Tanpa mempertimbangkan ketersediaan batas wilayah desa dalam bentuk peta yang ditetapkan oleh bupati/walikota.
- 5) **Kode 4 (empat)** merupakan desa yang tidak ada BPD tetapi ada kantor kepala desa di dalam wilayah desa. Tanpa mempertimbangkan ketersediaan batas wilayah desa dalam bentuk peta yang ditetapkan oleh bupati/walikota.
- 6) **Kode 5 (lima)** merupakan desa yang ada BPD dan ada kantor kepala desa di dalam wilayah desa. Tanpa mempertimbangkan ketersediaan batas wilayah desa dalam bentuk peta yang ditetapkan oleh bupati/walikota.

Rekategorisasi:

- 1) Tidak terdapat BPD atau kantor kepala desa.
- 2) Terdapat BPD dan kantor kepala desa.

40. Otonomi desa mengacu pada sumber penerimaan desa (Pendapatan Asli Desa, Alokasi Dana Desa, dan bagi hasil/bantuan/hibah) yang dimiliki oleh desa.

Kategori dari Bappenas:

- 1) **Kode 0 (nol)** merupakan desa yang tidak mempunyai sumber penerimaan desa (Pendapatan Asli Desa, Alokasi Dana Desa, dan bagi hasil/bantuan/hibah).
- 2) **Kode 1 (satu)** merupakan desa yang mempunyai rasio Pendapatan Asli Desa (PAD) terhadap total penerimaan desa kurang dari 0,06812184397134 (7 persen).
- 3) **Kode 2 (dua)** merupakan desa yang mempunyai rasio Pendapatan Asli Desa (PAD) terhadap total penerimaan desa lebih dari atau sama dengan 0,06812184397134 (7 persen) dan kurang dari 0,13624368794269 (14 persen).
- 4) **Kode 3 (tiga)** merupakan desa yang mempunyai rasio Pendapatan Asli Desa (PAD) terhadap total penerimaan desa lebih dari atau sama dengan 0,13624368794269 (14 persen) dan kurang dari 0,20436553191403 (20 persen).
- 5) **Kode 4 (empat)** merupakan desa yang mempunyai rasio Pendapatan Asli Desa (PAD) terhadap total penerimaan desa lebih dari atau sama dengan 0,20436553191403 (20 persen) dan kurang dari 0,27248737588537 (27 persen).
- 6) **Kode 5 (lima)** merupakan desa yang mempunyai rasio Pendapatan Asli Desa (PAD) terhadap total penerimaan desa lebih dari atau sama dengan 0,27248737588537 (27 persen).

Rekategorisasi:

- 1) Rasio PAD terhadap total penerimaan desa kurang dari 27 persen.
 - 2) Rasio PAD terhadap total penerimaan desa lebih dari atau sama dengan 27 persen.
41. Aset desa adalah barang milik desa yang berasal dari kekayaan asli desa, dibeli atau diperoleh atas beban Anggaran Pendapatan dan Belanja Desa atau perolehan hak lainnya yang sah (Undang-Undang Nomor 6 Tahun 2014 Tentang Desa).

Kategori dari Bappenas:

- 1) **Kode 0 (nol)** merupakan desa yang sama sekali tidak mempunyai kelima jenis aset/pendapatan desa yaitu: Pendapatan Asli Desa (PAD), tanah kas

desa/ulayat, bangunan desa (kantor kepala desa, balai desa, dan lain-lain), pasar desa (pasar hewan, pelelangan ikan, pelelangan hasil pertanian, dan lain-lain), maupun aset desa lainnya.

- 2) **Kode 1 (satu)** merupakan desa yang mempunyai 1 (satu) dari kelima jenis aset/pendapatan desa yaitu: Pendapatan Asli Desa (PAD), tanah kas desa/ulayat, bangunan desa (kantor kepala desa, balai desa, dan lain-lain), pasar desa (pasar hewan, pelelangan ikan, pelelangan hasil pertanian, dan lain-lain), maupun aset desa lainnya.
- 3) **Kode 2 (dua)** merupakan desa yang mempunyai 2 (dua) dari kelima jenis aset/pendapatan desa yaitu: Pendapatan Asli Desa (PAD), tanah kas desa/ulayat, bangunan desa (kantor kepala desa, balai desa, dan lain-lain), pasar desa (pasar hewan, pelelangan ikan, pelelangan hasil pertanian, dan lain-lain), maupun aset desa lainnya.
- 4) **Kode 3 (tiga)** merupakan desa yang mempunyai 3 (tiga) dari kelima jenis aset/pendapatan desa yaitu: Pendapatan Asli Desa (PAD), tanah kas desa/ulayat, bangunan desa (kantor kepala desa, balai desa, dan lain-lain), pasar desa (pasar hewan, pelelangan ikan, pelelangan hasil pertanian, dan lain-lain), maupun aset desa lainnya.
- 5) **Kode 4 (empat)** merupakan desa yang mempunyai 4 (empat) dari kelima jenis aset/pendapatan desa yaitu: Pendapatan Asli Desa (PAD), tanah kas desa/ulayat, bangunan desa (kantor kepala desa, balai desa, dan lain-lain), pasar desa (pasar hewan, pelelangan ikan, pelelangan hasil pertanian, dan lain-lain), maupun aset desa lainnya.
- 6) **Kode 5 (lima)** merupakan desa yang mempunyai 5 (lima) jenis aset/pendapatan desa yaitu: Pendapatan Asli Desa (PAD), tanah kas desa/ulayat, bangunan desa (kantor kepala desa, balai desa, dan lain-lain), pasar desa (pasar hewan, pelelangan ikan, pelelangan hasil pertanian, dan lain-lain), maupun aset desa lainnya.

Rekategorisasi:

- 1) Memiliki kurang dari atau sama dengan dua jenis aset desa.
- 2) Memiliki lebih dari dua jenis aset desa.

42. Kualitas SDM Kepala Desa mengacu pada pendidikan tertinggi yang ditamatkan.

Kategori dari Bappenas:

- 1) **Kode 0 (nol)** merupakan desa yang tidak mempunyai kepala desa.
- 2) **Kode 1 (satu)** merupakan desa yang mempunyai kepala desa dengan pendidikan tertinggi tidak tamat SD sederajat atau tidak pernah sekolah.
- 3) **Kode 2 (dua)** merupakan desa yang mempunyai kepala desa dengan pendidikan tertinggi yang ditamatkannya adalah SD sederajat.
- 4) **Kode 3 (tiga)** merupakan desa yang mempunyai kepala desa dengan pendidikan tertinggi yang ditamatkannya adalah SMP sederajat.
- 5) **Kode 4 (empat)** merupakan desa yang mempunyai kepala desa dengan pendidikan tertinggi yang ditamatkannya adalah SMU sederajat.
- 6) **Kode 5 (lima)** merupakan desa yang mempunyai kepala desa dengan pendidikan tertinggi yang ditamatkannya adalah akademi/DIII hingga S3.

Rekategorisasi:

- 1) Pendidikan kurang dari atau sama dengan SMU sederajat.
- 2) Pendidikan lebih dari SMU sederajat.

43. Kualitas SDM Sekretaris Desa mengacu pada pendidikan tertinggi yang ditamatkan.

Kategori dari Bappenas:

- 1) **Kode 0 (nol)** merupakan desa yang tidak mempunyai sekretaris desa.
- 2) **Kode 1 (satu)** merupakan desa yang mempunyai sekretaris desa dengan pendidikan tertinggi tidak tamat SD sederajat atau tidak pernah sekolah.
- 3) **Kode 2 (dua)** merupakan desa yang mempunyai sekretaris desa dengan pendidikan tertinggi yang ditamatkannya adalah SD sederajat atau SMP sederajat.
- 4) **Kode 3 (tiga)** merupakan desa yang mempunyai sekretaris desa dengan pendidikan tertinggi yang ditamatkannya adalah SMA sederajat.
- 5) **Kode 4 (empat)** merupakan desa yang mempunyai sekretaris desa dengan pendidikan tertinggi yang ditamatkannya adalah Akademi/DIII.

- 6) **Kode 5 (lima)** merupakan desa yang mempunyai sekretaris desa dengan pendidikan tertinggi yang ditamatkannya adalah Diploma IV/S1 hingga S3.

Rekategorisasi:

- 1) Pendidikan kurang dari atau sama dengan SMU sederajat
- 2) Pendidikan lebih dari SMU sederajat

3.2.3 Langkah-langkah Penerapan *Rare Event Weighted Logistic Regression*

Langkah-langkah analisis untuk menerapkan *RE-WLR* pada klasifikasi *imbalanced data* sebagai berikut:

1. Langkah-langkah menerapkan *RE-WLR* untuk klasifikasi desa tertinggal di Provinsi Jawa Timur.
 - 1) Persiapan Data
 - a. Imputasi data
 - b. Recoding data
 - 2) Analisis Deskriptif.
 - 3) Membagi data menjadi data *training* dan data *testing*.
 - 4) Mendapatkan penaksir $\hat{\beta}$, $B(\hat{\beta})$, $\tilde{\beta}$, \tilde{p}_i dengan *RE-WLR* pada data *training*
 - a. Menghitung nilai proporsi desa tertinggal di Jawa Timur (τ).
 - b. Menghitung nilai proporsi desa tertinggal dalam sampel (\bar{y}).
 - c. Menghitung probabilitas \hat{p}_i (2.2).
 - d. Menghitung varians v_i (2.11).
 - e. Menentukan pembobot w_i pada persamaan (2.19).
 - f. Menghitung *Adjusted response* (2.35).
 - g. Menghitung matriks kovarian (**Q**) pada persamaan (2.28).
 - h. Menentukan nilai Q_{ii} .
 - i. Menghitung *bias response* pada persamaan (2.27).
 - j. Menentukan matriks pembobot **D** dari pada persamaan (2.20).
 - k. Menghitung penaksir $\hat{\beta}$ dengan algoritma 4.
 - l. Menghitung bias($\hat{\beta}$) pada persamaan (2.22) dengan algoritma 5.

- m. Menghitung penaksir dengan bias terkoreksi ($\tilde{\beta}$) pada persamaan (2.30).
- n. Menghitung probabilitas optimal (\tilde{p}_i) dengan persamaan (2.37).
- 5) Mendapatkan model klasifikasi *RE-WLR* dari hasil langkah (4).
- 6) Memasukkan data *testing* ke dalam model klasifikasi *RE-WLR*, sehingga didapatkan hasil klasifikasi *RE-WLR*.
- 7) Evaluasi performansi *RE-WLR*
 - a. Membuat *confusion matrix* dari hasil klasifikasi dengan *RE-WLR*.
 - b. Menghitung ketepatan atau akurasi klasifikasi dari metode *RE-WLR* (2.38).
 - c. Menghitung *sensitivity* klasifikasi dari metode *RE-WLR* (2.39).
 - d. Menghitung *specificity* klasifikasi dari metode *RE-WLR* (2.40).
 - e. Menghitung nilai *AUC* dari metode *RE-WLR*.
 - f. Menghitung nilai *G-mean* dari metode *RE-WLR*.
- 2. Langkah-langkah untuk membandingkan tingkat ketepatan klasifikasi antara *TR-IRLS* dan *RE-WLR*.
 - 1) Menerapkan *TR-IRLS* untuk klasifikasi desa tertinggal di Provinsi Jawa Timur
 - i. Mendapatkan penaksir $\hat{\beta}$ dengan algoritma *TR-IRLS* pada data *training*.
 - a. Menghitung probabilitas \hat{p}_i (2.1).
 - b. Menghitung varians v_i (2.11).
 - c. Menghitung *Adjusted response* (2.14).
 - d. Menentukan matriks \mathbf{V} (2.12).
 - e. Menghitung penaksir $\hat{\beta}$ dengan algoritma 2.
 - ii. Mendapatkan model klasifikasi *TR-IRLS*
 - iii. Memasukkan data *testing* ke dalam model klasifikasi *TR-IRLS*, sehingga didapatkan hasil klasifikasi *TR-IRLS*
 - iv. Evaluasi performansi *TR-IRLS*
 - a. Membuat *confusion matrix* dari hasil klasifikasi dengan *TR-IRLS*.

- b. Menghitung ketepatan atau akurasi klasifikasi dari model *TR-IRLS* (2.38).
 - c. Menghitung *sensitivity* klasifikasi dari model *TR-IRLS* (2.39).
 - d. Menghitung *specificity* klasifikasi dari model *TR-IRLS* (2.40).
 - e. Menghitung nilai *AUC* dari model *TR-IRLS*.
 - f. Menghitung nilai *G-mean* dari model *TR-IRLS*.
- 2) Membandingkan performansi klasifikasi desa tertinggal di Provinsi Jawa Timur antara model *TR-IRLS* dan *RE-WLR*.
- a. Menentukan λ optimal pada model *TR-IRLS* dan *RE-WLR*.
 - b. Membandingkan ukuran ketepatan klasifikasi antara model *TR-IRLS* dan *RE-WLR* pada λ optimal.

BAB 4

HASIL DAN PEMBAHASAN

Bab ini terdiri dari tiga sub bab. Sub bab pertama membahas tentang penaksiran parameter *Rare Event Weighted Logistic Regression*. Sub bab kedua membahas tentang penerapan *Rare Event Weighted Logistic Regression* pada klasifikasi *imbalanced data* dengan studi kasus yang digunakan yaitu klasifikasi desa tertinggal di provinsi Jawa Timur. Sub bab ketiga membahas tentang perbandingan ketepatan klasifikasi antara *Rare Event Weighted Logistic Regression* dengan *Truncated Regularized Iteratively Reweighted Least Square*.

4.1 Penaksiran Parameter Model *Rare Event Weighted Logistic Regression* (RE-WLR)

Pada subbab ini akan dibahas mengenai penaksiran parameter model *Rare Event Weighted Logistic Regression* (RE-WLR). Terdapat dua metode umum untuk estimasi parameter yaitu *Least-Squares Estimation* (LSE) dan *Maximum Likelihood Estimation* (MLE). Metode penaksiran parameter yang biasa digunakan dalam regresi logistik adalah metode MLE. Variabel respon y memiliki sebaran Bernoulli dengan fungsi sebaran peluangnya adalah:

$$P(\mathbf{x}_i, y_i | \boldsymbol{\beta}) = \begin{cases} p_i & , \text{if } y = 1 \\ 1 - p_i & , \text{if } y = 0 \end{cases}$$

$$P(\mathbf{x}_i, y_i | \boldsymbol{\beta}) = (p_i)^{y_i} (1 - p_i)^{(1-y_i)}.$$

RE-WLR menerapkan regularisasi, *weighting*, dan *bias correction* pada Regresi Logistik. Langkah-langkah untuk mendapatkan penaksir parameter pada RE-WLR sebagai berikut:

a. *Regularized Logistic Regression*

1. Menentukan fungsi kemungkinan (*likelihood function*) pada model *Logistic Regression*.

Menurut Hosmer dan Lemeshow (1989), fungsi *likelihood* distribusi Bernoulli untuk n sampel bebas adalah

$$L(\boldsymbol{\beta}) = \prod_{i=1}^n (p_i)^{y_i} (1 - p_i)^{(1-y_i)} = \prod_{i=1}^n \left(\frac{e^{\mathbf{x}_i^T \boldsymbol{\beta}}}{1 + e^{\mathbf{x}_i^T \boldsymbol{\beta}}} \right)^{y_i} \left(\frac{1}{1 + e^{\mathbf{x}_i^T \boldsymbol{\beta}}} \right)^{(1-y_i)}.$$

Untuk memudahkan mencari nilai $\hat{\boldsymbol{\beta}}$ yang memaksimumkan fungsi likelihood, digunakan bentuk logaritma natural dari fungsi likelihood, yang disebut sebagai fungsi *log likelihood*.

$$\ln L(\boldsymbol{\beta}) = \sum_{i=1}^n \left(y_i \ln \left(\frac{e^{\mathbf{x}_i^T \boldsymbol{\beta}}}{1 + e^{\mathbf{x}_i^T \boldsymbol{\beta}}} \right) + (1 - y_i) \ln \left(\frac{1}{1 + e^{\mathbf{x}_i^T \boldsymbol{\beta}}} \right) \right).$$

$$\begin{aligned} \frac{\partial}{\partial \beta_j} \ln L(\boldsymbol{\beta}) &= \sum_{i=1}^n \left(y_i \left(\frac{x_{ij}}{1 + e^{\mathbf{x}_i^T \boldsymbol{\beta}}} \right) + (1 - y_i) \left(\frac{-x_{ij} e^{\mathbf{x}_i^T \boldsymbol{\beta}}}{1 + e^{\mathbf{x}_i^T \boldsymbol{\beta}}} \right) \right) \\ &= \sum_{i=1}^n \left(y_i x_{ij} \left(\frac{1}{1 + e^{\mathbf{x}_i^T \boldsymbol{\beta}}} \right) - (1 - y_i) x_{ij} \left(\frac{e^{\mathbf{x}_i^T \boldsymbol{\beta}}}{1 + e^{\mathbf{x}_i^T \boldsymbol{\beta}}} \right) \right) \\ &= \sum_{i=1}^n (y_i x_{ij} (1 - p_i) - (1 - y_i) x_{ij} p_i) \\ &= \sum_{i=1}^n (y_i x_{ij} - y_i x_{ij} p_i - x_{ij} p_i + y_i x_{ij} p_i) \\ &= \sum_{i=1}^n (x_{ij} (y_i - p_i)), \end{aligned}$$

dimana $j = 0 \dots k$ dan k adalah jumlah parameter.

Dalam bentuk matriks, turunan parsial pertama terhadap $\boldsymbol{\beta}$ dapat ditulis sebagai

$$\mathbf{X}^T (\mathbf{y} - \mathbf{p}) = \mathbf{0}.$$

Turunan parsial kedua terhadap $\boldsymbol{\beta}$ yaitu

$$\begin{aligned} \frac{\partial^2}{\partial \beta_j \partial \beta_k} \ln L(\boldsymbol{\beta}) &= \sum_{i=1}^n \left(\frac{-x_{ij} x_{ik} e^{\mathbf{x}_i^T \boldsymbol{\beta}}}{(1 + e^{\mathbf{x}_i^T \boldsymbol{\beta}})^2} \right) \\ &= \sum_{i=1}^n (-x_{ij} x_{ik} (p_i (1 - p_i))), \end{aligned}$$

dimana $v_i = p_i (1 - p_i)$ dan $\mathbf{V} = \text{diag}(v_1, \dots, v_n)$.

Matrik Hessian dapat ditulis sebagai

$$\mathbf{H}(\boldsymbol{\beta}) = -\mathbf{X}^T \mathbf{V} \mathbf{X}.$$

2. Membentuk *regularized log-likelihood function* yang diperoleh dengan menentukan *log-likelihood function* pada langkah (a.1) dan menambahkan parameter regularisasi $\frac{\lambda}{2} \|\boldsymbol{\beta}\|^2$

Fungsi *Regularized log likelihood* yang dibentuk yaitu

$$\begin{aligned} \ln L(\boldsymbol{\beta}) &= \sum_{i=1}^n \left(y_i \ln \left(\frac{e^{\mathbf{x}_i^T \boldsymbol{\beta}}}{1 + e^{\mathbf{x}_i^T \boldsymbol{\beta}}} \right) + (1 - y_i) \ln \left(\frac{1}{1 + e^{\mathbf{x}_i^T \boldsymbol{\beta}}} \right) \right) - \frac{\lambda}{2} \|\boldsymbol{\beta}\|^2 \\ \ln L(\boldsymbol{\beta}) &= \sum_{i=1}^n \left(y_i \left(\ln \left(\frac{e^{\mathbf{x}_i^T \boldsymbol{\beta}}}{1 + e^{\mathbf{x}_i^T \boldsymbol{\beta}}} \right) - \ln \left(\frac{1}{1 + e^{\mathbf{x}_i^T \boldsymbol{\beta}}} \right) \right) + \ln \left(\frac{1}{1 + e^{\mathbf{x}_i^T \boldsymbol{\beta}}} \right) \right) \\ &\quad - \frac{\lambda}{2} \|\boldsymbol{\beta}\|^2 \\ &= \sum_{i=1}^n \left(y_i \left(\ln(e^{\mathbf{x}_i^T \boldsymbol{\beta}}) - \ln(1 + e^{\mathbf{x}_i^T \boldsymbol{\beta}}) - \ln 1 + \ln(1 + e^{\mathbf{x}_i^T \boldsymbol{\beta}}) \right) \right. \\ &\quad \left. + \ln 1 - \ln(1 + e^{\mathbf{x}_i^T \boldsymbol{\beta}}) \right) - \frac{\lambda}{2} \|\boldsymbol{\beta}\|^2 \\ &= \sum_{i=1}^n \left(y_i (\ln(e^{\mathbf{x}_i^T \boldsymbol{\beta}}) - 0) + 0 - \ln(1 + e^{\mathbf{x}_i^T \boldsymbol{\beta}}) \right) - \frac{\lambda}{2} \|\boldsymbol{\beta}\|^2 \\ &= \sum_{i=1}^n \left(\ln(e^{\mathbf{x}_i^T \boldsymbol{\beta}})^{y_i} - \ln(1 + e^{\mathbf{x}_i^T \boldsymbol{\beta}}) \right) - \frac{\lambda}{2} \|\boldsymbol{\beta}\|^2 \\ &= \sum_{i=1}^n \ln \left(\frac{e^{y_i \mathbf{x}_i^T \boldsymbol{\beta}}}{1 + e^{\mathbf{x}_i^T \boldsymbol{\beta}}} \right) - \frac{\lambda}{2} \|\boldsymbol{\beta}\|^2 \\ &= - \sum_{i=1}^n \ln \left(e^{-y_i \mathbf{x}_i^T \boldsymbol{\beta}} (1 + e^{\mathbf{x}_i^T \boldsymbol{\beta}}) \right) - \frac{\lambda}{2} \|\boldsymbol{\beta}\|^2. \end{aligned}$$

b. *Regularized Weighted Logistic Regression*

1. Menentukan pembobot.

King dan Zheng merekomendasikan dua metode yaitu *prior correction* dan *weighting*. Pada *RE-WLR*, metode yang digunakan adalah *weighting*. Joint distribution dari \mathbf{y} dan \mathbf{X} dalam sampel yaitu

$$f_s(\mathbf{y}, \mathbf{X}|\boldsymbol{\beta}) = P_s(\mathbf{X}|\mathbf{y}, \boldsymbol{\beta})P_s(\mathbf{y})$$

dimana $\boldsymbol{\beta}$ adalah parameter yang tidak diketahui dan akan diestimasi. \mathbf{X} adalah matriks variabel prediktor, maka probabilitas bersyarat \mathbf{X} dalam sampel adalah sama dengan probabilitas bersyarat \mathbf{X} dalam populasi atau $P_s(\mathbf{X}|\mathbf{y}, \boldsymbol{\beta}) = P(\mathbf{X}|\mathbf{y}, \boldsymbol{\beta})$. Probabilitas bersyarat dalam populasi yaitu

$$P(\mathbf{X}|\mathbf{y}, \boldsymbol{\beta}) = \frac{f(\mathbf{y}, \mathbf{X}|\boldsymbol{\beta})}{P(\mathbf{y})},$$

$$\text{dan } f(\mathbf{y}, \mathbf{X}|\boldsymbol{\beta}) = P(\mathbf{y}|\mathbf{X}, \boldsymbol{\beta})P(\mathbf{X}).$$

$$\begin{aligned} f_s(\mathbf{y}, \mathbf{X}|\boldsymbol{\beta}) &= \frac{P_s(\mathbf{y})}{P(\mathbf{y})} P(\mathbf{y}|\mathbf{X}, \boldsymbol{\beta})P(\mathbf{X}) \\ &= \frac{H}{Q} P(\mathbf{y}|\mathbf{X}, \boldsymbol{\beta})P(\mathbf{X}). \end{aligned}$$

dimana $\frac{H}{Q} = \frac{P_s(\mathbf{y})}{P(\mathbf{y})}$, dengan H adalah proporsi y dalam sampel dan Q adalah proporsi y dalam populasi. Selanjutnya fungsi likelihood yaitu

$$L_{Endogenous} = \prod_{i=1}^n \frac{H_i}{Q_i} P(y_i|\mathbf{x}_i, \boldsymbol{\beta})P(\mathbf{x}_i),$$

dimana $\frac{H_i}{Q_i} = \left(\frac{\bar{y}}{\tau}\right) y_i + \left(\frac{1-\bar{y}}{1-\tau}\right) (1 - y_i)$, dengan \bar{y} adalah proporsi kejadian dalam sampel dan τ adalah penduga proporsi kejadian dalam populasi.

Fungsi *Likelihood* yaitu

$$\begin{aligned} \ln L(\boldsymbol{\beta}|\mathbf{y}, \mathbf{X}) &= \sum_{i=1}^n \frac{Q_i}{H_i} \ln P(y_i|\mathbf{x}_i, \boldsymbol{\beta}) \\ &= \sum_{i=1}^n \frac{Q_i}{H_i} \ln \left(\frac{e^{y_i \mathbf{x}_i^T \boldsymbol{\beta}}}{1 + e^{\mathbf{x}_i^T \boldsymbol{\beta}}} \right) \\ &= \sum_{i=1}^n w_i \ln \left(\frac{e^{y_i \mathbf{x}_i^T \boldsymbol{\beta}}}{1 + e^{\mathbf{x}_i^T \boldsymbol{\beta}}} \right), \end{aligned}$$

dimana $w_i = \frac{Q_i}{H_i}$. Maka pembobot $w_i = \left(\frac{\bar{y}}{\tau}\right) y_i + \left(\frac{1-\bar{y}}{1-\tau}\right) (1 - y_i)$.

Jika $y_i = 1$, maka $w_i = \frac{\tau}{\bar{y}} = w_1$ dan jika $y_i = 0$, maka $w_i = \frac{1-\tau}{1-\bar{y}} = w_0$.

2. Menentukan fungsi kemungkinan (*likelihood function*) pada model *Weighted Logistic Regression*.

Fungsi *likelihood* yang dibentuk yaitu

$$L_w(\boldsymbol{\beta}) = \prod_{i=1}^n (p_i)^{w_1 y_i} (1 - p_i)^{w_0 (1 - y_i)},$$

dimana $w_1 = \frac{\tau}{\bar{y}}$ dan $w_0 = \frac{1-\tau}{1-\bar{y}}$.

3. Membentuk *regularized weighted log-likelihood function* $\ln L_w(\boldsymbol{\beta})$ yang diperoleh dengan menentukan *log-likelihood function* pada langkah (b.2) dan menambahkan parameter regularisasi $\frac{\lambda}{2} \|\boldsymbol{\beta}\|^2$.

$$\begin{aligned} \ln L_w(\boldsymbol{\beta}) &= \sum_{i=1}^n w_i \ln \left(\frac{e^{y_i \mathbf{x}_i^T \boldsymbol{\beta}}}{1 + e^{\mathbf{x}_i^T \boldsymbol{\beta}}} \right) - \frac{\lambda}{2} \|\boldsymbol{\beta}\|^2 \\ &= \sum_{i=1}^n w_i \ln \left(\frac{e^{y_i \mathbf{x}_i^T \boldsymbol{\beta}}}{1 + e^{\mathbf{x}_i^T \boldsymbol{\beta}}} \right) - \frac{\lambda}{2} \left(\sqrt{\beta_0^2 + \beta_1^2 + \beta_2^2 + \dots + \beta_k^2} \right)^2 \\ &= \sum_{i=1}^n w_i \ln \left(\frac{e^{y_i \mathbf{x}_i^T \boldsymbol{\beta}}}{1 + e^{\mathbf{x}_i^T \boldsymbol{\beta}}} \right) - \frac{\lambda}{2} (\beta_0^2 + \beta_1^2 + \beta_2^2 + \dots + \beta_k^2). \end{aligned}$$

4. Mendapatkan vektor *gradient* dengan mencari turunan parsial pertama dari *regularized weighted log-likelihood function* (b.3), yaitu $\nabla_{\boldsymbol{\beta}} \ln L_w(\boldsymbol{\beta})$.

$$\mathbf{g}(\boldsymbol{\beta}) = \nabla_{\boldsymbol{\beta}} \ln L_w(\boldsymbol{\beta}) = \begin{bmatrix} \frac{\partial}{\partial \beta_0} \ln L_w(\boldsymbol{\beta}) \\ \frac{\partial}{\partial \beta_1} \ln L_w(\boldsymbol{\beta}) \\ \vdots \\ \frac{\partial}{\partial \beta_j} \ln L_w(\boldsymbol{\beta}) \\ \vdots \\ \frac{\partial}{\partial \beta_{42}} \ln L_w(\boldsymbol{\beta}) \end{bmatrix}$$

$$\begin{aligned} \frac{\partial}{\partial \beta_0} \ln L_w(\boldsymbol{\beta}) &= \sum_{i=1}^n w_i \left(\left(\frac{1 + e^{\mathbf{x}_i^T \boldsymbol{\beta}}}{e^{y_i \mathbf{x}_i^T \boldsymbol{\beta}}} \right) \left(\left(\frac{y_i e^{y_i \mathbf{x}_i^T \boldsymbol{\beta}}}{1 + e^{\mathbf{x}_i^T \boldsymbol{\beta}}} \right) + \left(\frac{e^{y_i \mathbf{x}_i^T \boldsymbol{\beta}} (-e^{\mathbf{x}_i^T \boldsymbol{\beta}})}{(1 + e^{\mathbf{x}_i^T \boldsymbol{\beta}})^2} \right) \right) \right) \\ &\quad - \frac{\lambda}{2} \cdot 2\beta_0 \end{aligned}$$

$$\begin{aligned}
&= \sum_{i=1}^n w_i \left(\left(\frac{1 + e^{\mathbf{x}_i^T \boldsymbol{\beta}}}{e^{y_i \mathbf{x}_i^T \boldsymbol{\beta}}} \right) \left(\frac{y_i e^{y_i \mathbf{x}_i^T \boldsymbol{\beta}} (1 + e^{\mathbf{x}_i^T \boldsymbol{\beta}})}{(1 + e^{\mathbf{x}_i^T \boldsymbol{\beta}})^2} \right) \right. \\
&\quad \left. + \left(\frac{e^{y_i \mathbf{x}_i^T \boldsymbol{\beta}} (-e^{\mathbf{x}_i^T \boldsymbol{\beta}})}{(1 + e^{\mathbf{x}_i^T \boldsymbol{\beta}})^2} \right) \right) - \lambda \beta_0 \\
&= \sum_{i=1}^n w_i \left(\left(\frac{1 + e^{\mathbf{x}_i^T \boldsymbol{\beta}}}{e^{y_i \mathbf{x}_i^T \boldsymbol{\beta}}} \right) \left(\frac{e^{y_i \mathbf{x}_i^T \boldsymbol{\beta}} (y_i (1 + e^{\mathbf{x}_i^T \boldsymbol{\beta}}) - e^{\mathbf{x}_i^T \boldsymbol{\beta}})}{(1 + e^{\mathbf{x}_i^T \boldsymbol{\beta}})^2} \right) \right) - \lambda \beta_0 \\
&= \sum_{i=1}^n w_i \left(\frac{(y_i (1 + e^{\mathbf{x}_i^T \boldsymbol{\beta}}) - e^{\mathbf{x}_i^T \boldsymbol{\beta}})}{1 + e^{\mathbf{x}_i^T \boldsymbol{\beta}}} \right) - \lambda \beta_0 \\
&= \sum_{i=1}^n w_i (y_i - p_i) - \lambda \beta_0. \\
\\
\frac{\partial}{\partial \beta_1} \ln L_w(\boldsymbol{\beta}) &= \sum_{i=1}^n w_i \left(\left(\frac{1 + e^{\mathbf{x}_i^T \boldsymbol{\beta}}}{e^{y_i \mathbf{x}_i^T \boldsymbol{\beta}}} \right) \left(\frac{y_i x_{i1} e^{y_i \mathbf{x}_i^T \boldsymbol{\beta}}}{1 + e^{\mathbf{x}_i^T \boldsymbol{\beta}}} \right) \right. \\
&\quad \left. + \left(\frac{e^{y_i \mathbf{x}_i^T \boldsymbol{\beta}} (-x_{i1} e^{\mathbf{x}_i^T \boldsymbol{\beta}})}{(1 + e^{\mathbf{x}_i^T \boldsymbol{\beta}})^2} \right) \right) - \frac{\lambda}{2} \cdot 2 \beta_1 \\
&= \sum_{i=1}^n w_i \left(\left(\frac{1 + e^{\mathbf{x}_i^T \boldsymbol{\beta}}}{e^{y_i \mathbf{x}_i^T \boldsymbol{\beta}}} \right) \left(\frac{y_i x_{i1} e^{y_i \mathbf{x}_i^T \boldsymbol{\beta}} (1 + e^{\mathbf{x}_i^T \boldsymbol{\beta}})}{(1 + e^{\mathbf{x}_i^T \boldsymbol{\beta}})^2} \right) \right. \\
&\quad \left. + \left(\frac{e^{y_i \mathbf{x}_i^T \boldsymbol{\beta}} (-x_{i1} e^{\mathbf{x}_i^T \boldsymbol{\beta}})}{(1 + e^{\mathbf{x}_i^T \boldsymbol{\beta}})^2} \right) \right) - \lambda \beta_1 \\
&= \sum_{i=1}^n w_i \left(\left(\frac{1 + e^{\mathbf{x}_i^T \boldsymbol{\beta}}}{e^{y_i \mathbf{x}_i^T \boldsymbol{\beta}}} \right) \left(\frac{x_{i1} e^{y_i \mathbf{x}_i^T \boldsymbol{\beta}} (y_i (1 + e^{\mathbf{x}_i^T \boldsymbol{\beta}}) - e^{\mathbf{x}_i^T \boldsymbol{\beta}})}{(1 + e^{\mathbf{x}_i^T \boldsymbol{\beta}})^2} \right) \right) \\
&\quad - \lambda \beta_1 \\
&= \sum_{i=1}^n w_i \left(x_{i1} \left(\frac{(y_i (1 + e^{\mathbf{x}_i^T \boldsymbol{\beta}}) - e^{\mathbf{x}_i^T \boldsymbol{\beta}})}{1 + e^{\mathbf{x}_i^T \boldsymbol{\beta}}} \right) \right) - \lambda \beta_1.
\end{aligned}$$

$$\frac{\partial}{\partial \beta_1} \ln L_w(\boldsymbol{\beta}) = \sum_{i=1}^n w_i x_{i1} (y_i - p_i) - \lambda \beta_1 .$$

⋮

$$\frac{\partial}{\partial \beta_j} \ln L_w(\boldsymbol{\beta}) = \sum_{i=1}^n w_i x_{ij} (y_i - p_i) - \lambda \beta_j .$$

⋮

$$\frac{\partial}{\partial \beta_{42}} \ln L_w(\boldsymbol{\beta}) = \sum_{i=1}^n w_i x_{i42} (y_i - p_i) - \lambda \beta_{42} .$$

Dalam bentuk matriks,

$$\nabla_{\boldsymbol{\beta}} \ln L_w(\boldsymbol{\beta}) = \mathbf{X}^T \mathbf{W}(\mathbf{y} - \mathbf{p}) - \lambda \boldsymbol{\beta},$$

dimana $\mathbf{W} = \text{diag}(w_1, \dots, w_n)$ dan \mathbf{p} adalah vektor probabilitas.

5. Mendapatkan matriks *Hessian* dengan mencari turunan parsial kedua dari *regularized weighted log-likelihood function* (b.3), yaitu $\nabla_{\boldsymbol{\beta}}^2 \ln L_w(\boldsymbol{\beta})$.

$$\mathbf{H}(\boldsymbol{\beta}) = \nabla_{\boldsymbol{\beta}}^2 \ln L_w(\boldsymbol{\beta})$$

$$\begin{aligned} &= \begin{bmatrix} \frac{\partial^2}{\partial \beta_0 \partial \beta_0} \ln L_w(\boldsymbol{\beta}) & \cdots & \frac{\partial^2}{\partial \beta_0 \partial \beta_k} \ln L_w(\boldsymbol{\beta}) & \cdots & \frac{\partial^2}{\partial \beta_0 \partial \beta_{42}} \ln L_w(\boldsymbol{\beta}) \\ \vdots & \ddots & \vdots & \cdots & \vdots \\ \frac{\partial^2}{\partial \beta_j \partial \beta_0} \ln L_w(\boldsymbol{\beta}) & \cdots & \frac{\partial^2}{\partial \beta_j \partial \beta_k} \ln L_w(\boldsymbol{\beta}) & \cdots & \frac{\partial^2}{\partial \beta_j \partial \beta_{42}} \ln L_w(\boldsymbol{\beta}) \\ \vdots & \cdots & \vdots & \ddots & \vdots \\ \frac{\partial^2}{\partial \beta_{42} \partial \beta_0} \ln L_w(\boldsymbol{\beta}) & \cdots & \frac{\partial^2}{\partial \beta_{42} \partial \beta_k} \ln L_w(\boldsymbol{\beta}) & \cdots & \frac{\partial^2}{\partial \beta_{42} \partial \beta_{42}} \ln L_w(\boldsymbol{\beta}) \end{bmatrix} \\ \\ \frac{\partial^2}{\partial \beta_j \partial \beta_k} \ln L_w(\boldsymbol{\beta}) &= \sum_{i=1}^n -w_i x_{ij} \left(\left(\frac{x_{ik} e^{\mathbf{x}_i^T \boldsymbol{\beta}}}{1 + e^{\mathbf{x}_i^T \boldsymbol{\beta}}} \right) + \left(\frac{x_{ik} e^{\mathbf{x}_i^T \boldsymbol{\beta}} (-e^{\mathbf{x}_i^T \boldsymbol{\beta}})}{(1 + e^{\mathbf{x}_i^T \boldsymbol{\beta}})^2} \right) \right) - \lambda \\ &= \sum_{i=1}^n -w_i x_{ij} \left(\frac{x_{ik} e^{\mathbf{x}_i^T \boldsymbol{\beta}} (1 + e^{\mathbf{x}_i^T \boldsymbol{\beta}} - e^{\mathbf{x}_i^T \boldsymbol{\beta}})}{(1 + e^{\mathbf{x}_i^T \boldsymbol{\beta}})(1 + e^{\mathbf{x}_i^T \boldsymbol{\beta}})} \right) - \lambda \\ &= \sum_{i=1}^n -w_i x_{ij} \left(\frac{x_{ik} (1 + e^{\mathbf{x}_i^T \boldsymbol{\beta}} - e^{\mathbf{x}_i^T \boldsymbol{\beta}})}{(1 + e^{\mathbf{x}_i^T \boldsymbol{\beta}})} \right) - \lambda \\ &= \sum_{i=1}^n -w_i x_{ij} x_{ik} \left(\frac{e^{\mathbf{x}_i^T \boldsymbol{\beta}}}{1 + e^{\mathbf{x}_i^T \boldsymbol{\beta}}} \right) \left(\frac{1 - e^{\mathbf{x}_i^T \boldsymbol{\beta}}}{1 + e^{\mathbf{x}_i^T \boldsymbol{\beta}}} \right) - \lambda \end{aligned}$$

$$= \sum_{i=1}^n \left(-x_{ij} x_{ik} w_i (p_i (1 - p_i)) \right) - \lambda$$

Dalam bentuk matriks

$$\nabla_{\beta}^2 \ln L_w(\beta) = -\mathbf{X}^T \mathbf{D} \mathbf{X} - \lambda \mathbf{I},$$

dimana $\mathbf{D} = \text{diag} (v_i w_i)$.

6. Mendapatkan iterasi Newton-Raphson untuk $\hat{\beta}$ menggunakan $\nabla_{\beta} \ln L_w(\beta)$ dan $\nabla_{\beta}^2 \ln L_w(\beta)$ dari langkah (b.4) dan (b.5).

$$\begin{aligned} \hat{\beta}^{(c+1)} &= \hat{\beta}^{(c)} - (\nabla_{\beta}^2 \ln L_w(\beta))^{-1} (\nabla_{\beta} \ln L_w(\beta)) \\ &= \hat{\beta}^{(c)} + (\mathbf{X}^T \mathbf{D} \mathbf{X} + \lambda \mathbf{I})^{-1} (\mathbf{X}^T \mathbf{W}(\mathbf{y} - \mathbf{p}) - \lambda \hat{\beta}^{(c)}). \end{aligned}$$

Jika $\hat{\beta}^{(c)} = (\mathbf{X}^T \mathbf{D} \mathbf{X} + \lambda \mathbf{I})^{-1} (\mathbf{X}^T \mathbf{D} \mathbf{X} + \lambda \mathbf{I}) \hat{\beta}^{(c)}$, maka

$$\begin{aligned} \hat{\beta}^{(c+1)} &= (\mathbf{X}^T \mathbf{D} \mathbf{X} + \lambda \mathbf{I})^{-1} (\mathbf{X}^T \mathbf{D} \mathbf{X} + \lambda \mathbf{I}) \hat{\beta}^{(c)} \\ &\quad + (\mathbf{X}^T \mathbf{D} \mathbf{X} + \lambda \mathbf{I})^{-1} (\mathbf{X}^T \mathbf{W}(\mathbf{y} - \mathbf{p}) - \lambda \hat{\beta}^{(c)}) \\ &= (\mathbf{X}^T \mathbf{D} \mathbf{X} + \lambda \mathbf{I})^{-1} \mathbf{X}^T \mathbf{D} (\mathbf{X} \hat{\beta}^{(c)} + (\mathbf{X}^T)^{-1} \mathbf{D}^{-1} \lambda \hat{\beta}^{(c)} + \mathbf{D}^{-1} \mathbf{W}(\mathbf{y} - \mathbf{p}) \\ &\quad - (\mathbf{X}^T)^{-1} \mathbf{D}^{-1} \lambda \hat{\beta}^{(c)}) \\ &= (\mathbf{X}^T \mathbf{D} \mathbf{X} + \lambda \mathbf{I})^{-1} \mathbf{X}^T \mathbf{D} (\mathbf{X} \hat{\beta}^{(c)} + \mathbf{D}^{-1} \mathbf{W}(\mathbf{y} - \mathbf{p})) \\ &= (\mathbf{X}^T \mathbf{D} \mathbf{X} + \lambda \mathbf{I})^{-1} \mathbf{X}^T \mathbf{D} (\mathbf{X} \hat{\beta}^{(c)} + \mathbf{V}^{-1} \mathbf{W}^{-1} \mathbf{W}(\mathbf{y} - \mathbf{p})) \\ &= (\mathbf{X}^T \mathbf{D} \mathbf{X} + \lambda \mathbf{I})^{-1} \mathbf{X}^T \mathbf{D} (\mathbf{X} \hat{\beta}^{(c)} + \mathbf{V}^{-1} (\mathbf{y} - \mathbf{p})). \end{aligned}$$

Update Newton-Raphson untuk β pada iterasi ke (c+1) yaitu

$$\hat{\beta}^{(c+1)} = (\mathbf{X}^T \mathbf{D} \mathbf{X} + \lambda \mathbf{I})^{-1} \mathbf{X}^T \mathbf{D} \mathbf{z}^{(c)},$$

dimana $\mathbf{z}^{(c)} = \mathbf{X} \hat{\beta}^{(c)} + \mathbf{V}^{-1} (\mathbf{y} - \mathbf{p})$.

7. Mendapatkan penaksir parameter $\hat{\beta}$ dengan menggunakan metode Truncated Newton untuk subproblem *WLS* dengan algoritma *CG* linier.
8. Menentukan vektor bias $\mathbf{B}(\hat{\beta})$.

$$\eta_i = \mathbf{x}_i^T \beta, \text{ sehingga } p_i = \frac{1}{1+e^{-\eta_i}}, \text{ dan } (1 - p_i) = \frac{e^{-\eta_i}}{1+e^{-\eta_i}}.$$

$$p_i^* = \left(\frac{1}{1 + e^{-\eta_i}} \right)^{w_1} = p_i^{w_1}.$$

$$p_i^{*'} = \frac{\partial p_i^*}{\partial \eta_i} = w_1 \left(\frac{1}{1 + e^{-\eta_i}} \right)^{w_1-1} (-1) (1 + e^{-\eta_i})^{-2} (-e^{-\eta_i})$$

$$\begin{aligned}
&= w_1 \left(\frac{1}{1 + e^{-\eta_i}} \right)^{w_1} (1 + e^{-\eta_i}) (1 + e^{-\eta_i})^{-2} (e^{-\eta_i}) \\
&= w_1 \left(\frac{1}{1 + e^{-\eta_i}} \right)^{w_1} \left(\frac{e^{-\eta_i}}{1 + e^{-\eta_i}} \right) \\
&= w_1 p_i^{w_1} (1 - p_i). \\
p_i^{*''} &= \frac{\partial^2 p_i^*}{\partial \eta_i^2} = w_1^2 \left(\frac{1}{1 + e^{-\eta_i}} \right)^{w_1} \left(\frac{e^{-\eta_i}}{1 + e^{-\eta_i}} \right)^2 \\
&\quad + w_1 \left(\frac{1}{1 + e^{-\eta_i}} \right)^{w_1} \left(\frac{-e^{-\eta_i}}{1 + e^{-\eta_i}} + \frac{e^{-\eta_i}(-e^{-\eta_i})}{(1 + e^{-\eta_i})^2} \right) \\
&= w_1^2 \left(\frac{1}{1 + e^{-\eta_i}} \right)^{w_1} \left(\frac{e^{-\eta_i}}{1 + e^{-\eta_i}} \right)^2 \\
&\quad + w_1 \left(\frac{1}{1 + e^{-\eta_i}} \right)^{w_1} \left(\frac{e^{-\eta_i}(-1 - e^{-\eta_i} + e^{-\eta_i})}{(1 + e^{-\eta_i})^2} \right) \\
&= w_1^2 \left(\frac{1}{1 + e^{-\eta_i}} \right)^{w_1} \left(\frac{e^{-\eta_i}}{1 + e^{-\eta_i}} \right)^2 \\
&\quad + w_1 \left(\frac{1}{1 + e^{-\eta_i}} \right)^{w_1} \left(\left(\frac{e^{-\eta_i}}{1 + e^{-\eta_i}} \right) \left(-\frac{1}{1 + e^{-\eta_i}} \right) \right) \\
&= w_1 p_i^{w_1} (1 - p_i) w_1 (1 - p_i) + w_1 p_i^{w_1} (1 - p_i) (-p_i) \\
&= w_1 p_i^{w_1} (1 - p_i) (w_1 - w_1 p_i) + w_1 p_i^{w_1} (1 - p_i) (-p_i) \\
&= w_1 p_i^{w_1} (1 - p_i) (w_1 - (1 + w_1) p_i)
\end{aligned}$$

Vektor Bias untuk WLR adalah

$$\mathbf{B}(\hat{\boldsymbol{\beta}}) = (\mathbf{X}^T \mathbf{D} \mathbf{X})^{-1} \mathbf{X}^T \mathbf{D} \boldsymbol{\xi},$$

dimana elemen ke-i dari vector $\boldsymbol{\xi}$ adalah

$$\begin{aligned}
\xi_i &= -0,5 \left(\frac{p_i^{*''}}{p_i^{*'}} \right) Q_{ii} = 0,5 Q_{ii} ((1 + w_1) p_i - w_1), \\
\xi_i &= -0,5 \left(\frac{w_1 p_i^{w_1} (1 - p_i) (w_1 - (1 + w_1) p_i)}{w_1 p_i^{w_1} (1 - p_i)} \right) Q_{ii} \\
\xi_i &= 0,5 Q_{ii} ((1 + w_1) p_i - w_1)
\end{aligned}$$

dengan Q_{ii} adalah elemen diagonal dari matriks kovarians \mathbf{Q} dimana sekarang menjadi $\mathbf{Q} = \mathbf{X}(\mathbf{X}^T \mathbf{D} \mathbf{X})^{-1} \mathbf{X}^T$.

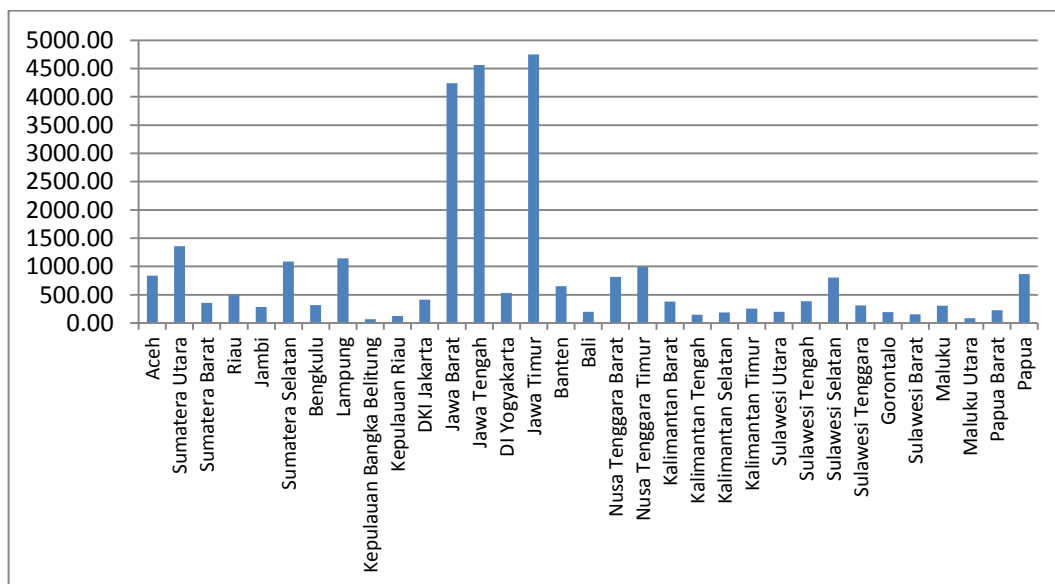
9. Mendapatkan penaksir parameter $\tilde{\beta}$ dengan bias terkoreksi.

$$\tilde{\beta} = \hat{\beta} - \mathbf{B}(\hat{\beta}).$$

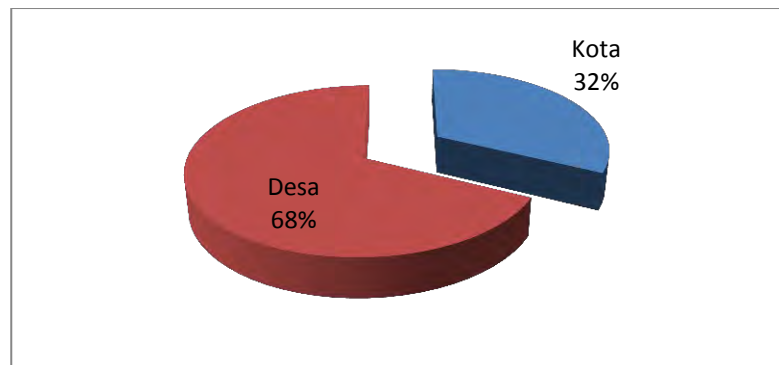
4.2 Klasifikasi Desa Tertinggal di Provinsi Jawa Timur

4.2.1 Gambaran Umum Provinsi Jawa Timur Menurut Variabel Penelitian

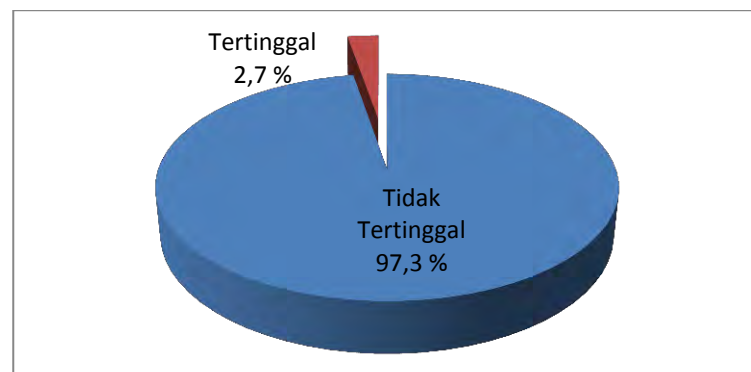
Pada tahun 2014, Jawa Timur memiliki jumlah penduduk miskin terbesar di Indonesia yaitu sebesar 4,75 juta jiwa. Berdasarkan lokasi tempat tinggal, penduduk miskin digolongkan menjadi dua yaitu penduduk miskin perkotaan dan pedesaan. Sebagian besar penduduk miskin di Jawa Timur tinggal di daerah pedesaan yaitu sebanyak 3,22 juta jiwa. Hal tersebut menunjukkan bahwa kantong kemiskinan di Jawa Timur terdapat di wilayah pedesaan. Lampiran 3 dan Gambar 4.3 menunjukkan sebanyak 208 (2,7%) desa termasuk dalam klasifikasi desa tertinggal dan 7.513 (97,3%) termasuk dalam klasifikasi desa tidak tertinggal.



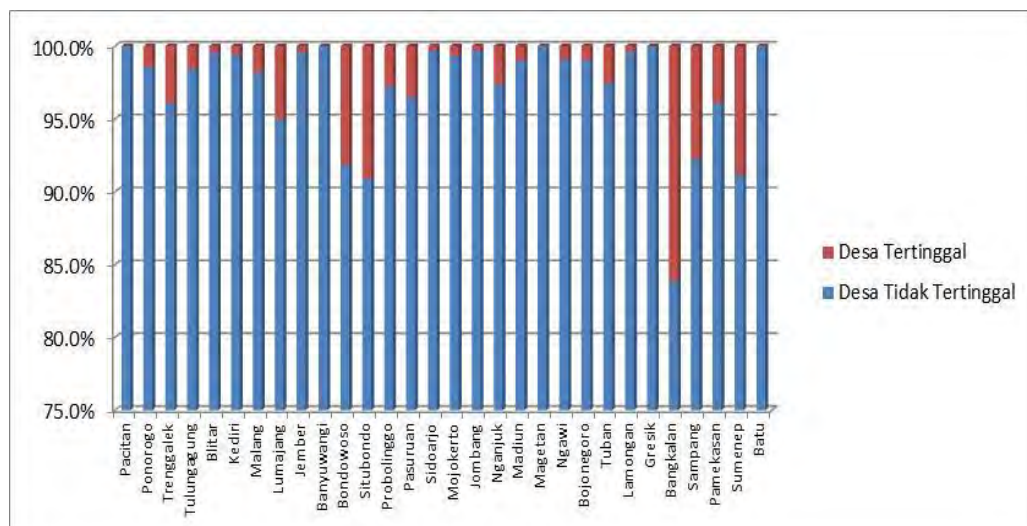
Gambar 4.1 Jumlah Penduduk Miskin Menurut Provinsi di Indonesia Tahun 2014 (dalam 000). (Sumber : BPS)



Gambar 4.2 Persentase Penduduk Miskin Menurut Lokasi Tempat Tinggal di Provinsi Jawa Timur Tahun 2014. (Sumber : BPS)



Gambar 4.3 Persentase Desa Tertinggal dan desa Tidak Tertinggal di Provinsi Jawa Timur Tahun 2014. (Sumber : Bappenas)



Gambar 4.4 Persentase Desa Tertinggal dan Desa Tidak Tertinggal Menurut Kabupaten/Kota di Provinsi Jawa Timur Tahun 2014. (Sumber : Bappenas)

Salah satu hal yang menunjukkan tingkat pembangunan desa pada masing-masing kabupaten/kota adalah jumlah dan persentase desa tertinggal di wilayah tersebut. Berdasarkan Lampiran 3 dan Gambar 4.4, kabupaten di Pulau Madura dan daerah tapal kuda memiliki persentase desa tertinggal yang lebih tinggi daripada kabupaten/kota lain di Jawa Timur, yaitu Kabupaten Bangkalan sebesar 16,1% (44 desa), Kabupaten Sumenep sebesar 8,8% (29 desa), Kabupaten Sampang sebesar 7,8% (14 desa), Kabupaten Situbondo sebesar 9,1% (12 desa), Kabupaten Bondowoso sebesar 8,1% (17 desa), dan Kabupaten Lumajang sebesar 5,1% (10 desa). Beberapa kabupaten/kota tidak memiliki desa tertinggal atau 100% desanya termasuk desa tidak tertinggal, diantaranya Kabupaten Pacitan, Kabupaten Banyuwangi, Kabupaten Magetan, Kabupaten Gresik, dan Kota Batu. Gambaran kondisi tersebut menunjukkan bahwa terdapat kesenjangan pembangunan pada beberapa daerah di kabupaten dan kota di Provinsi Jawa Timur.

4.2.2 Klasifikasi Desa Tertinggal di Provinsi Jawa Timur dengan RE-WLR

Pembentukan model dilakukan dengan membagi data menjadi data *training* dan data *testing*. Data *training* digunakan untuk membangun model dan data *testing* untuk validasi model. Penelitian ini menggunakan lima jenis partisi data yaitu partisi data 50:50 (50% data sebagai data *training* dan sisa 50% data sebagai data *testing*), 60:40, 70:30, 80:20, dan 90:10. Pembagian data *training* dan *testing* dilakukan secara random dan stratifikasi.

Pada penelitian ini, nilai parameter λ yang digunakan adalah 1 sampai dengan 10, jadi nilai $\tilde{\beta}$ yang didapatkan akan bergantung pada data masing-masing partisi dan nilai lambda. Selanjutnya nilai $\tilde{\beta}$ yang diperoleh pada tiap model dimasukkan ke dalam persamaan

$$\tilde{p}_i = \frac{\exp(\tilde{\beta}_0 + \tilde{\beta}_1 x_{i1} + \tilde{\beta}_2 x_{i2} + \dots + \tilde{\beta}_{42} x_{i42})}{1 + \exp(\tilde{\beta}_0 + \tilde{\beta}_1 x_{i1} + \tilde{\beta}_2 x_{i2} + \dots + \tilde{\beta}_{42} x_{i42})},$$

dimana \tilde{p}_i adalah probabilitas suatu desa untuk masuk kategori desa tertinggal. Jika $\tilde{p}_i < 0,5$ maka desa diprediksi untuk masuk kelas desa tidak

tertinggal (0), dan jika $\tilde{p}_i \geq 0,5$ maka desa diprediksi untuk masuk kelas desa tertinggal (1).

a. Partisi Data Random

Tabel 4.1 menunjukkan jumlah data kelas positif (1) dan kelas negatif (0) yang masuk ke dalam data *training* dan *testing* dengan pembagian data yang dilakukan secara random. Proporsi kelas positif (*rare event*) tidak sama untuk data *training* dan data *testing*. Data aktual dan data hasil prediksi dari model klasifikasi *RE-WLR* disajikan dalam *confusion matrix* pada Lampiran 7 (data *training*) dan Lampiran 8 (data *testing*). *Confusion matrix* menunjukkan jumlah desa tertinggal yang diprediksi tepat sebagai desa tertinggal dan diprediksi tidak tepat sebagai desa tidak tertinggal, serta jumlah desa tidak tertinggal yang diprediksi tepat sebagai desa tidak tertinggal dan diprediksi tidak tepat sebagai desa tertinggal.

Tabel 4.1 Data Jumlah Kelas Positif dan Negatif pada Data *Training* dan Data *Testing*

Partisi Data	<i>Training</i>				<i>Testing</i>			
	0	1	Jumlah	% kelas positif	0	1	Jumlah	% kelas positif
50:50	3759	101	3860	2,6	3754	107	3861	2,8
60:40	4491	142	4633	3,1	3022	66	3088	2,1
70:30	5252	153	5405	2,8	2261	55	2316	2,4
80:20	6009	168	6177	2,7	1504	40	1544	2,6
90:10	6761	188	6949	2,7	752	20	772	2,6

Kinerja klasifikasi model *RE-WLR* dapat dilihat dari ukuran ketepatan klasifikasi pada data *testing*. Lampiran 10 menyajikan secara lengkap nilai *accuracy*, *sensitiftiy*, *specificity*, *AUC*, dan *G-mean* masing-masing model yang diperoleh pada partisi data dan λ yang berbeda. Nilai total akurasi (*accuracy*) menunjukkan tingkat ketepatan klasifikasi model secara keseluruhan yang dinyatakan dalam persen. Setiap nilai parameter λ pada masing-masing partisi data memberikan nilai $\tilde{\beta}$ dan total akurasi (*accuracy*) yang berbeda. Model *RE-WLR* yang memberikan nilai total akurasi tertinggi yaitu pada model partisi data 60:40 dengan λ sebesar 2 dan 3 yaitu 98,48%. Sedangkan rata-rata total akurasi tertinggi *RE-WLR* diperoleh dengan $\lambda = 2$, yaitu sebesar 98,06%.

Tabel 4.2 Rata-rata Kinerja Klasifikasi *RE-WLR* untuk Klasifikasi Desa Tertinggal di Jawa Timur

λ	Ukuran Ketepatan Klasifikasi				
	<i>Acc</i>	<i>Sens</i>	<i>Spec</i>	<i>AUC</i>	<i>g-mean</i>
1	98,04%	42,75%	99,45%	0,7110	65,19%
2	98,06%	39,27%	99,57%	0,6942	62,49%
3	98,00%	36,06%	99,59%	0,6783	59,88%
4	98,02%	33,21%	99,69%	0,6645	57,37%
5	97,97%	28,66%	99,75%	0,6420	53,38%
6	98,01%	28,83%	99,79%	0,6431	53,52%
7	97,98%	26,98%	99,81%	0,6339	51,73%
8	97,99%	26,48%	99,83%	0,6316	51,20%
9	97,97%	25,63%	99,83%	0,6273	50,38%
10	97,92%	23,00%	99,84%	0,6142	47,69%
Rata-rata	98,00%	31,09%	99,71%	0,6540	55,28%
Median	98,06%	30,15%	99,73%	0,6504	54,89%

Sensitivity menunjukkan tingkat ketepatan klasifikasi model untuk memprediksi kelas positif. Nilai *sensitivity* tertinggi diperoleh pada model partisi 80:20 dengan $\lambda = 1$ yaitu sebesar 45%. Secara umum, rata-rata nilai *sensitivity* yang menunjukkan ketepatan *RE-WLR* untuk memprediksi kelas positif (*rare event*) atau desa tertinggal dengan benar yaitu sebesar 31,09%.

Specificity menunjukkan tingkat ketepatan klasifikasi model untuk memprediksi kelas negatif. Nilai *specificity* tertinggi diperoleh pada model partisi 90:10 dengan λ sebesar 6, 7, 8, 9, dan 10 yaitu sebesar 100%. yang berarti model *RE-WLR* pada partisi data dan λ tersebut dapat memprediksi desa tidak tertinggal dengan benar seluruhnya. Secara umum, rata-rata nilai *specificity* yang menunjukkan ketepatan *RE-WLR* untuk memprediksi kelas positif atau desa tidak tertinggal dengan benar yaitu sebesar 99,71%.

Kinerja klasifikasi *imbalanced data* dapat dilihat dari nilai *AUC* dan *G-mean*. Model klasifikasi terbaik yang memiliki nilai *AUC* dan *G-mean* paling tinggi dibandingkan model lainnya yaitu model *RE-WLR* dengan nilai $\tilde{\beta}$ yang didapatkan dari partisi data 80:20 dengan λ sebesar 1. Nilai $\tilde{\beta}$ dari model tersebut dapat dilihat di Lampiran 6c. Probabilitas suatu desa untuk masuk kategori desa tertinggal yaitu

$$\begin{aligned}\tilde{p}_i &= \frac{\exp(\tilde{\beta}_0 + \tilde{\beta}_1 x_{i1} + \tilde{\beta}_2 x_{i2} + \dots + \tilde{\beta}_{42} x_{i42})}{1 + \exp(\tilde{\beta}_0 + \tilde{\beta}_1 x_{i1} + \tilde{\beta}_2 x_{i2} + \dots + \tilde{\beta}_{42} x_{i42})} \\ &= \frac{\exp(-15,94 + 1,18 x_{i1} - 0,11 x_{i2} + \dots + 1,16 x_{i42})}{1 + \exp(-15,94 + 1,18 x_{i1} - 0,11 x_{i2} + \dots + 1,16 x_{i42})}.\end{aligned}$$

Dari nilai $\tilde{\beta}_j$ yang diperoleh, dapat ditunjukkan urutan variabel yang signifikan dalam memprediksi suatu desa untuk masuk ke dalam kategori desa tertinggal (Lampiran 19). Sepuluh variabel yang berada di urutan atas adalah ketersediaan puskesmas (X7), kelengkapan pemerintahan desa (X38), bahan bakar untuk memasak (X20), aksesibilitas jalan (X27), kualitas fasilitas komunikasi seluler (X24), kondisi penerangan di jalan utama (X19), biaya per kilometer transportasi ke kantor camat (X31), ketersediaan SMA sederajat (X4), sumber air untuk minum (X21), dan ketersediaan TK/RA/BA (X1).

Dari seluruh model klasifikasi *RE-WLR* yang didapatkan pada penelitian ini, didapatkan nilai optimal λ yaitu 1, dimana rata-rata nilai *sensitify*, *AUC*, dan *G-mean* lebih tinggi dari nilai rata-rata nilai *sensitify*, *AUC*, dan *G-mean* pada λ selainnya.

Tabel 4.3 Perbandingan Kinerja Klasifikasi *RE-WLR* pada Data *Training* dan Data *Testing*

λ	Training					Testing				
	Acc	Sens	Spec	AUC	g-mean	Acc	Sens	Spec	AUC	g-mean
1	98,11%	46,12%	99,61%	0,7286	67,76%	98,04%	42,75%	99,45%	0,7110	65,19%
2	98,03%	41,33%	99,66%	0,7050	64,15%	98,06%	39,27%	99,57%	0,6942	62,49%
3	98,00%	38,02%	99,72%	0,6887	61,55%	98,00%	36,06%	99,59%	0,6783	59,88%
4	97,88%	32,59%	99,75%	0,6617	56,85%	98,02%	33,21%	99,69%	0,6645	57,37%
5	97,86%	30,65%	99,79%	0,6522	55,20%	97,97%	28,66%	99,75%	0,6420	53,38%
6	97,83%	29,00%	99,81%	0,6440	53,74%	98,01%	28,83%	99,79%	0,6431	53,52%
7	97,80%	27,71%	99,81%	0,6376	52,52%	97,98%	26,98%	99,81%	0,6339	51,73%
8	97,75%	25,33%	99,83%	0,6258	50,22%	97,99%	26,48%	99,83%	0,6316	51,20%
9	97,72%	24,11%	99,83%	0,6197	48,99%	97,97%	25,63%	99,83%	0,6273	50,38%
10	97,66%	21,92%	99,83%	0,6088	46,68%	97,92%	23,00%	99,84%	0,6142	47,69%
Rata-rata	97,86%	31,68%	99,76%	0,6572	55,77%	98,00%	31,09%	99,71%	0,6540	55,28%
Median	97,88%	29,48%	99,78%	0,6465	54,25%	98,06%	30,15%	99,73%	0,6504	54,89%

Tujuan regularisasi pada penelitian ini adalah untuk memperoleh generalisasi yang lebih baik atau mengatasi *over fitting*. Suatu model klasifikasi dinyatakan tidak *over fitting* apabila hasil model klasifikasi memberikan ketepatan yang relatif sama antara data *training* dan data *testing*. Tabel 4.3 menunjukkan perbandingan kinerja klasifikasi *RE-WLR* pada data *training* dan data *testing*.

Ukuran ketepatan klasifikasi yang terdiri dari *accuracy*, *sensitivity*, *specificity*, *AUC*, dan *G-mean* pada data *training* maupun data *testing* menunjukkan nilai yang relatif sama, sehingga dapat dinyatakan bahwa tidak terjadi *over fitting* pada model *RE-WLR* untuk klasifikasi desa tertinggal di Provinsi Jawa Timur.

b. Partisi Data Stratifikasi

Pembentukan data *training* dan data *testing* dilakukan dengan stratifikasi. Data dipisahkan menjadi data kelas positif (1) dan data kelas negatif (0). Setiap kelas diambil sejumlah data berdasarkan proporsi partisi data yang akan dibentuk. Misalkan pada pembagian data partisi 80:20, pada kelas positif 80% data diambil sebagai data positif *training* dan sisanya 20% sebagai data positif *testing*. Pada kelas negatif, 80% data diambil sebagai data negatif *training* dan sisanya 20% sebagai data negatif *testing*. Selanjutnya, 80% data positif *training* dan 80% data negatif *training* digabung menjadi data 80% *training*. Sedangkan 20% data positif *testing* dan 20% data negatif *testing* digabung menjadi data 20% *testing*. Cara pembagian data dengan stratifikasi akan membentuk partisi data dengan proporsi kelas yang sama antara data *training* dan data *testing*. Tabel 4.4 menunjukkan jumlah data kelas positif (1) dan kelas negatif (0) yang masuk ke dalam data *training* dan *testing* dengan stratifikasi.

Tabel 4.4 Data Jumlah Kelas Positif dan Negatif pada Data *Training* dan Data *Testing* Stratifikasi

Partisi Data	<i>Training</i>				<i>Testing</i>			
	0	1	Jumlah	% kelas positif	0	1	Jumlah	% kelas positif
50:50	3756	104	3860	2,7	3757	104	3861	2,7
60:40	4508	125	4633	2,7	3005	83	3088	2,7
70:30	5259	146	5405	2,7	2254	62	2316	2,7
80:20	6010	166	6176	2,7	1503	42	1545	2,7
90:10	6762	187	6949	2,7	751	21	772	2,7

Data aktual dan data hasil prediksi dari model klasifikasi *RE-WLR* dengan stratifikasi disajikan dalam *confusion matrix* pada Lampiran 11 (data *training*) dan Lampiran 12 (data *testing*). *Confusion matrix* menunjukkan jumlah desa tertinggal yang diprediksi tepat sebagai desa tertinggal dan diprediksi tidak tepat sebagai

desa tidak tertinggal, serta jumlah desa tidak tertinggal yang diprediksi tepat sebagai desa tidak tertinggal dan diprediksi tidak tepat sebagai desa tertinggal.

Tabel 4.5 Rata-rata Kinerja Klasifikasi *RE-WLR* (dengan Stratifikasi) untuk Klasifikasi Desa Tertinggal di Jawa Timur

λ	Ukuran Ketepatan Klasifikasi				
	<i>Acc</i>	<i>Sens</i>	<i>Spec</i>	<i>AUC</i>	<i>g-mean</i>
1	97,98%	40,93%	99,56%	0,7025	63,79%
2	97,88%	35,79%	99,60%	0,6769	59,54%
3	97,93%	38,36%	99,58%	0,6897	61,66%
4	97,70%	26,86%	99,67%	0,6327	51,60%
5	97,70%	25,98%	99,69%	0,6284	50,74%
6	97,84%	33,59%	99,62%	0,6660	57,47%
7	97,59%	20,02%	99,75%	0,5988	44,41%
8	97,57%	19,97%	99,72%	0,5985	44,43%
9	97,55%	19,35%	99,72%	0,5954	43,70%
10	97,51%	17,12%	99,74%	0,5843	41,02%
Rata-rata	97,70%	26,22%	99,68%	0,6295	50,38%
Median	97,67%	24,70%	99,67%	0,6222	49,63%

Kinerja klasifikasi model *RE-WLR* dapat dilihat dari ukuran ketepatan klasifikasi pada data *testing*. Lampiran 14 menyajikan secara lengkap nilai *accuracy*, *sensitify*, *specificity*, *AUC*, dan *G-mean* masing-masing model yang diperoleh pada partisi data dan λ yang berbeda pada data *testing* stratifikasi. Nilai total akurasi (*accuracy*) menunjukkan tingkat ketepatan klasifikasi model secara keseluruhan yang dinyatakan dalam persen. Setiap nilai parameter λ pada masing-masing partisi data memberikan nilai $\tilde{\beta}$ dan total akurasi (*accuracy*) yang berbeda. Model *RE-WLR* yang memberikan nilai total akurasi tertinggi yaitu pada model partisi data 50:50 dengan λ sebesar 1 yaitu 98,14%. Sedangkan rata-rata total akurasi tertinggi *RE-WLR* diperoleh dengan $\lambda = 1$, yaitu sebesar 97,98%.

Sensitify menunjukkan tingkat ketepatan klasifikasi model untuk memprediksi kelas positif. Nilai *sensitify* tertinggi diperoleh pada model partisi 50:50 dengan $\lambda = 1$ yaitu sebesar 45,19%. Rata-rata nilai *sensitify* tertinggi yang menunjukkan ketepatan *RE-WLR* untuk memprediksi kelas positif (*rare event*) atau desa tertinggal dengan benar yaitu sebesar 40,93% pada $\lambda = 1$.

Specificity menunjukkan tingkat ketepatan klasifikasi model untuk memprediksi kelas negatif. Nilai *specificity* tertinggi diperoleh pada model partisi 50:50 dengan λ sebesar 4 dan 10 yaitu sebesar 99,87%. Secara umum, rata-rata nilai *specificity* yang menunjukkan ketepatan *RE-WLR* untuk memprediksi kelas positif atau desa tidak tertinggal dengan benar yaitu sebesar 99,68%.

Model klasifikasi terbaik pada data stratifikasi yang memiliki nilai *AUC* dan *G-mean* paling tinggi dibandingkan model lainnya yaitu model *RE-WLR* dengan nilai $\tilde{\beta}$ yang didapatkan dari partisi data 50:50 dengan λ sebesar 1. Nilai $\tilde{\beta}$ dari model tersebut dapat dilihat di Lampiran 6d. Probabilitas suatu desa untuk masuk kategori desa tertinggal yaitu

$$\begin{aligned}\tilde{p}_i &= \frac{\exp(\tilde{\beta}_0 + \tilde{\beta}_1 x_{i1} + \tilde{\beta}_2 x_{i2} + \dots + \tilde{\beta}_{42} x_{i42})}{1 + \exp(\tilde{\beta}_0 + \tilde{\beta}_1 x_{i1} + \tilde{\beta}_2 x_{i2} + \dots + \tilde{\beta}_{42} x_{i42})} \\ &= \frac{\exp(-14,11 + 1,00 x_{i1} - 0,03 x_{i2} + \dots + 1,27 x_{i42})}{1 + \exp(-14,11 + 1,00 x_{i1} - 0,03 x_{i2} + \dots + 1,27 x_{i42})}.\end{aligned}$$

Dari seluruh model klasifikasi *RE-WLR* yang didapatkan dengan pembagian data stratifikasi, didapatkan nilai optimal λ yaitu 1, dimana rata-rata nilai *sensitivity*, *AUC*, dan *G-mean* lebih tinggi dari nilai rata-rata nilai *sensitivity*, *AUC*, dan *G-mean* pada λ selainnya.

Tabel 4.6 Perbandingan Kinerja Klasifikasi *RE-WLR* pada Data *Training* dan Data *Testing* Stratifikasi

λ	Training					Testing				
	Acc	Sens	Spec	AUC	g-mean	Acc	Sens	Spec	AUC	g-mean
1	98,17%	45,99%	99,61%	0,7280	67,68%	97,98%	40,93%	99,56%	0,7025	63,79%
2	98,08%	39,86%	99,69%	0,6977	62,98%	97,88%	35,79%	99,60%	0,6769	59,54%
3	98,05%	38,15%	99,71%	0,6893	61,61%	97,93%	38,36%	99,58%	0,6897	61,66%
4	97,94%	32,27%	99,76%	0,6602	56,51%	97,70%	26,86%	99,67%	0,6327	51,60%
5	97,93%	31,11%	99,78%	0,6544	55,59%	97,70%	25,98%	99,69%	0,6284	50,74%
6	97,88%	28,47%	99,80%	0,6413	53,24%	97,84%	33,59%	99,62%	0,6660	57,47%
7	97,85%	26,75%	99,82%	0,6328	51,55%	97,59%	20,02%	99,75%	0,5988	44,41%
8	97,84%	26,30%	99,82%	0,6306	51,11%	97,57%	19,97%	99,72%	0,5985	44,43%
9	97,81%	24,83%	99,83%	0,6233	49,60%	97,55%	19,35%	99,72%	0,5954	43,70%
10	97,78%	23,38%	99,84%	0,6161	48,08%	97,51%	17,12%	99,74%	0,5843	41,02%
Rata-rata	97,93%	31,71%	99,77%	0,6574	55,79%	97,70%	26,22%	99,68%	0,6295	50,38%
Median	97,96%	30,67%	99,79%	0,6525	55,33%	97,67%	24,70%	99,67%	0,6222	49,63%

Suatu model klasifikasi dinyatakan tidak *over fitting* apabila hasil model klasifikasi memberikan ketepatan yang relatif sama antara data *training* dan data *testing*. Tabel 4.6 menunjukkan perbandingan kinerja klasifikasi *RE-WLR* pada data *training* dan data *testing*. Ukuran ketepatan klasifikasi yang terdiri dari *accuracy*, *sensitiftity*, *specificity*, *AUC*, dan *G-mean* pada data *training* maupun data *testing* menunjukkan nilai yang relatif sama, sehingga dapat dinyatakan bahwa tidak terjadi *over fitting* pada model *RE-WLR* (dengan stratifikasi) untuk klasifikasi desa tertinggal di Provinsi Jawa Timur.

4.3 Perbandingan Ketepatan Klasifikasi Desa Tertinggal antara Model *TR-IRLS* dan *RE-WLR*

a. Partisi Data Random

Penerapan *TR-IRLS* dilakukan pada data yang sama dengan *RE-WLR* (Tabel 4.1). Confusion *matrix* untuk model *TR-IRLS* dapat dilihat di Lampiran 7 (data *training*) dan lampiran 8 (data *testing*). Kinerja model klasifikasi dapat dilihat dari ukuran ketepatan klasifikasi pada data *testing*. Hasil perbandingan *Accuracy*, *Sensitivity*, *Specificity*, *AUC*, dan *G-mean* hasil klasifikasi *TR-IRLS* dan *RE-WLR* dapat dilihat di Lampiran 10.

Jika dibandingkan antara model *TR-IRLS* dan *RE-WLR* pada λ optimal, rata-rata *sensitiftity*, *AUC* dan *G-mean* *RE-WLR* lebih tinggi dari *TR-IRLS* (Tabel 4.7). Secara deskriptif, sebaran nilai-nilai ukuran ketepatan klasifikasi tersebut dapat dilihat pada perbandingan Box Plot *TR-IRLS* dan *RE-WLR* (Lampiran 15). Tabel 4.7 dan Lampiran 15 menunjukkan bahwa pada λ optimal secara deskriptif *RE-WLR* memiliki performansi klasifikasi yang lebih baik daripada *TR-IRLS*.

Tabel 4.7 Perbandingan Kinerja Klasifikasi *TR-IRLS* dan *RE-WLR* pada λ Optimal

$\lambda=1$	<i>TR-IRLS</i>					<i>RE-WLR</i>				
	Acc	Sens	Spec	AUC	g-mean	Acc	Sens	Spec	AUC	g-mean
50:50	97,80%	40,19%	99,44%	0,6981	63,22%	97,82%	42,99%	99,39%	0,7119	65,37%
60:40	98,38%	46,97%	99,50%	0,7324	68,36%	98,25%	43,94%	99,44%	0,7169	66,10%
70:30	98,14%	40,00%	99,56%	0,6978	63,11%	98,14%	41,82%	99,51%	0,7067	64,51%
80:20	97,80%	37,50%	99,40%	0,6845	61,05%	97,93%	45,00%	99,34%	0,7217	66,86%
90:10	98,06%	40,00%	99,60%	0,6980	63,12%	98,06%	40,00%	99,60%	0,6980	63,12%
Rata-rata	98,04%	40,93%	99,50%	0,7022	63,77%	98,04%	42,75%	99,45%	0,7110	65,19%

Selanjutnya, uji statistik untuk perbedaan nilai ketepatan klasifikasi dilakukan dengan uji peringkat bertanda Wilcoxon.

Uji peringkat bertanda Wilcoxon

i. Perbandingan nilai *Sensitifity RE-WLR* dan *TR-IRLS*

H_0 : Tidak ada perbedaan yang signifikan antara *sensitifity RE-WLR* dan *TR-IRLS*.

H_1 : Terdapat perbedaan yang signifikan antara *sensitifity RE-WLR* dan *TR-IRLS*.

Pada uji ini, α yang digunakan adalah 0,05. Dari hasil uji peringkat bertanda Wilcoxon (Lampiran 16), didapatkan nilai $p\text{-value} > \alpha$ sehingga gagal tolak H_0 yang berarti bahwa tidak terdapat perbedaan yang signifikan antara *sensitifity RE-WLR* dan *TR-IRLS*.

ii. Perbandingan nilai *AUC RE-WLR* dan *TR-IRLS*

H_0 : Tidak ada perbedaan yang signifikan antara nilai *AUC RE-WLR* dan *TR-IRLS*.

H_1 : Terdapat perbedaan yang signifikan antara nilai *AUC RE-WLR* dan *TR-IRLS*.

Pada uji ini, α yang digunakan adalah 0,05. Dari hasil uji peringkat bertanda Wilcoxon (Lampiran 16), didapatkan nilai $p\text{-value} > \alpha$ sehingga gagal tolak H_0 yang berarti bahwa tidak terdapat perbedaan yang signifikan antara nilai *AUC RE-WLR* dan *TR-IRLS*.

iii. Perbandingan nilai *G-mean RE-WLR* dan *TR-IRLS*

H_0 : Tidak ada perbedaan yang signifikan antara *G-mean RE-WLR* dan *TR-IRLS*.

H_1 : Terdapat perbedaan yang signifikan antara *G-mean RE-WLR* dan *TR-IRLS*.

Pada uji ini, α yang digunakan adalah 0,05. Dari hasil uji peringkat bertanda Wilcoxon (Lampiran 16), didapatkan nilai $p\text{-value} > \alpha$ sehingga gagal tolak H_0 yang berarti bahwa tidak terdapat perbedaan yang signifikan antara nilai *G-mean RE-WLR* dan *TR-IRLS*.

b. Partisi Data Stratifikasi

Penerapan *TR-IRLS* dilakukan pada data stratifikasi yang sama dengan *RE-WLR* (Tabel 4.4). *Confusion matrix* untuk model *TR-IRLS* dapat dilihat di Lampiran 11 (data *training*) dan lampiran 12 (data *testing*). Kinerja model klasifikasi dapat dilihat dari ukuran ketepatan klasifikasi pada data *testing*. Hasil perbandingan *Accuracy*, *Sensitivity*, *Specificity*, *AUC*, dan *G-mean* hasil klasifikasi *TR-IRLS* dan *RE-WLR* dengan data stratifikasi dapat dilihat di Lampiran 14.

Jika dibandingkan antara model *TR-IRLS* dan *RE-WLR* pada λ optimal, rata-rata *sensitivity*, *AUC* dan *G-mean* *RE-WLR* lebih tinggi dari *TR-IRLS* (Tabel 4.8) Secara deskriptif, sebaran nilai-nilai ukuran ketepatan klasifikasi tersebut dapat dilihat pada perbandingan Box Plot *TR-IRLS* dan *RE-WLR* (Lampiran 17). Tabel 4.8 dan Lampiran 17 menunjukkan bahwa pada λ optimal secara deskriptif *RE-WLR* memiliki performansi klasifikasi yang lebih baik daripada *TR-IRLS*.

Tabel 4.8 Perbandingan Kinerja Klasifikasi *TR-IRLS* dan *RE-WLR* pada λ Optimal (Data Stratifikasi)

$\lambda=1$	<i>TR-IRLS</i>					<i>RE-WLR</i>				
	Acc	Sens	Spec	AUC	g-mean	Acc	Sens	Spec	AUC	g-mean
50:50	98,14%	45,19%	99,60%	0,7240	67,09%	98,14%	45,19%	99,60%	0,7240	67,09%
60:40	97,99%	44,58%	99,47%	0,7202	66,59%	98,02%	44,58%	99,50%	0,7204	66,60%
70:30	97,88%	35,48%	99,60%	0,6754	59,45%	97,93%	38,71%	99,56%	0,6913	62,08%
80:20	98,19%	35,71%	99,93%	0,6782	59,74%	97,86%	38,10%	99,53%	0,6881	61,58%
90:10	97,80%	38,10%	99,47%	0,6878	61,56%	97,93%	38,10%	99,60%	0,6885	61,60%
Rata-rata	98,00%	39,81%	99,61%	0,6971	62,89%	97,98%	40,93%	99,56%	0,7025	63,79%

Selanjutnya, uji statistik untuk perbedaan nilai ketepatan klasifikasi dilakukan dengan uji peringkat bertanda Wilcoxon.

Uji peringkat bertanda Wilcoxon

i. Perbandingan nilai *Sensitivity* *RE-WLR* dan *TR-IRLS*

H_0 : Tidak ada perbedaan yang signifikan antara *sensitivity* *RE-WLR* dan *TR-IRLS*.

H_1 : Terdapat perbedaan yang signifikan antara *sensitivity* *RE-WLR* dan *TR-IRLS*.

Pada uji ini α yang digunakan adalah 0,05. Dari hasil uji peringkat bertanda Wilcoxon (Lampiran 18), didapatkan nilai $p\text{-value} > \alpha$ sehingga gagal tolak H_0 yang berarti bahwa tidak terdapat perbedaan yang signifikan antara *sensitifity RE-WLR* dan *TR-IRLS*.

ii. Perbandingan nilai *AUC RE-WLR* dan *TR-IRLS*

H_0 : Tidak ada perbedaan yang signifikan antara nilai *AUC RE-WLR* dan *TR-IRLS*.

H_1 : Terdapat perbedaan yang signifikan antara nilai *AUC RE-WLR* dan *TR-IRLS*.

Pada uji ini α yang digunakan adalah 0,05. Dari hasil uji peringkat bertanda Wilcoxon (Lampiran 18), didapatkan nilai $p\text{-value} > \alpha$ sehingga gagal tolak H_0 yang berarti bahwa tidak terdapat perbedaan yang signifikan antara nilai *AUC RE-WLR* dan *TR-IRLS*.

iii. Perbandingan nilai *G-mean RE-WLR* dan *TR-IRLS*

H_0 : Tidak ada perbedaan yang signifikan antara *G-mean RE-WLR* dan *TR-IRLS*.

H_1 : Terdapat perbedaan yang signifikan antara *G-mean RE-WLR* dan *TR-IRLS*.

Pada uji ini α yang digunakan adalah 0,05. Dari hasil uji peringkat bertanda Wilcoxon (Lampiran 18), didapatkan nilai $p\text{-value} > \alpha$ sehingga gagal tolak H_0 yang berarti bahwa tidak terdapat perbedaan yang signifikan antara nilai *G-mean RE-WLR* dan *TR-IRLS*.

BAB 5

KESIMPULAN DAN SARAN

5.1 Kesimpulan

Berdasarkan analisis data dan pembahasan diperoleh kesimpulan sebagai berikut:

1. Penaksiran parameter model *RE-WLR* dilakukan dengan menggunakan metode *Maximum Likelihood Estimation* (MLE). Estimasi parameter tersebut menghasilkan persamaan yang tidak *closed form* sehingga dilanjutkan dengan metode iterasi *Newton Raphson*. Terdapat problem numerik yaitu dalam mendapatkan invers dari matriks Hessian berukuran besar yang mengakibatkan pemrosesan menjadi lama, serta kekonvergenan pada IRLS, sehingga digunakan pendekatan metode *Truncated Newton* dengan algoritma *Linear CG*.
2. Pada penelitian ini, pembentukan partisi data untuk data *training* dan data *testing* dilakukan dengan dua cara yaitu dengan random dan stratifikasi. Nilai regularisasi λ yang digunakan adalah 1 sampai dengan 10.
 - a. Pada partisi data random, nilai optimal λ model klasifikasi *RE-WLR* yaitu 1. Dengan nilai optimal λ tersebut, rata-rata nilai *accuracy* model *RE-WLR* untuk klasifikasi desa tertinggal di Jawa Timur yaitu 98,04%, rata-rata *sensitivity* yaitu 42,75%, rata-rata *specificity* yaitu 99,45%, rata-rata *AUC* yaitu 0,7110, dan rata-rata *G-mean* yaitu 65,19%. Nilai *sensitivity* menunjukkan ketepatan *RE-WLR* untuk memprediksi kelas positif (*rare event*) atau desa tertinggal.
 - b. Pada partisi data stratifikasi, nilai optimal λ model klasifikasi *RE-WLR* yaitu 1. Dengan nilai optimal λ tersebut, rata-rata nilai *accuracy* model *RE-WLR* untuk klasifikasi desa tertinggal di Jawa Timur yaitu 97,98%, rata-rata *sensitivity* yaitu 40,93%, rata-rata *specificity* yaitu 99,56%, rata-rata *AUC* yaitu 0,7025, dan rata-rata *G-mean* yaitu 63,79%.
3. Perbandingan model klasifikasi *TR-IRLS* dan *RE-WLR* pada studi kasus klasifikasi desa tertinggal di Jawa Timur dilakukan pada partisi data random

dan stratifikasi. Pada partisi data random maupun stratifikasi, perbandingan *TR-IRLS* dan *RE-WLR* pada nilai optimal λ sama-sama menunjukkan bahwa secara deskriptif *RE-WLR* memiliki performansi klasifikasi yang lebih baik daripada *TR-IRLS*, namun dengan perbedaan yang tidak signifikan. Rata-rata nilai *sensitifty* *RE-WLR* yang lebih tinggi daripada *TR-IRLS* menunjukkan bahwa *RE-WLR* bisa memprediksi kelas minoritas (*rare event*) atau desa tertinggal dengan lebih baik dibandingkan *TR-IRLS*.

5.2 Saran

Beberapa saran yang dapat diberikan berdasarkan hasil yang telah diperoleh adalah sebagai berikut:

1. Metode yang digunakan dalam penelitian ini masih terbatas pada kasus klasifikasi biner. Penelitian selanjutnya bisa mencoba penggunaan *RE-WLR* pada kasus klasifikasi *multiclass*.
2. Pada penelitian ini kasus yang digunakan adalah klasifikasi desa tertinggal di Provinsi Jawa Timur Tahun 2014. Hasil yang diperoleh tidak dapat disimpulkan secara *general*. Maka dari itu, penelitian selanjutnya dapat menggunakan data simulasi. Penggunaan *RE-WLR* pada data simulasi akan memberikan hasil yang lebih optimal dan kesimpulan secara *general*.
3. Pada penelitian ini, algoritma *RE-WLR* diterapkan pada data variabel prediktor kategori. Penelitian selanjutnya bisa menerapkan *RE-WLR* pada data variabel prediktor kontinyu.

Lampiran 1. Data Indeks Pembangunan Desa (IPD) Jawa Timur 2014

No	Kode Wilayah	Kabupaten	Kecamatan	Desa	IPD
1	3501010001	PACITAN	DONOROJO	WIDORO	53.92
2	3501010002	PACITAN	DONOROJO	SAWAHAN	56.01
3	3501010003	PACITAN	DONOROJO	KALAK	69.45
4	3501010004	PACITAN	DONOROJO	SENDANG	55.33
5	3501010005	PACITAN	DONOROJO	KLEPU	61.88
6	3501010006	PACITAN	DONOROJO	GEDOMPOL	60.13
7	3501010007	PACITAN	DONOROJO	CEMENG	55.47
8	3501010008	PACITAN	DONOROJO	GENDARAN	64.39
9	3501010009	PACITAN	DONOROJO	SUKODONO	60.07
10	3501010010	PACITAN	DONOROJO	SEKAR	60.89
11	3501010011	PACITAN	DONOROJO	DONOROJO	76.92
12	3501010012	PACITAN	DONOROJO	BELAH	68.37
13	3501020001	PACITAN	PUNUNG	BOMO	71.59
14	3501020002	PACITAN	PUNUNG	SOOKA	69.95
15	3501020003	PACITAN	PUNUNG	PUNUNG	82.18
16	3501020004	PACITAN	PUNUNG	MENDOLO KIDUL	62.56
17	3501020005	PACITAN	PUNUNG	MENDOLO LOR	60.29
18	3501020006	PACITAN	PUNUNG	KENDAL	66.49
19	3501020007	PACITAN	PUNUNG	PITON	62.45
20	3501020008	PACITAN	PUNUNG	WARENG	66.13
21	3501020009	PACITAN	PUNUNG	MANTREN	68.42
22	3501020010	PACITAN	PUNUNG	PLOSO	68.55
23	3501020011	PACITAN	PUNUNG	GONDOSARI	73.26
24	3501020012	PACITAN	PUNUNG	TINATAR	64.38
25	3501020013	PACITAN	PUNUNG	KEBONSARI	56.22
26	3501030001	PACITAN	PRINGKUKU	WATU KARUNG	67.61
27	3501030002	PACITAN	PRINGKUKU	DERSONO	63.10
28	3501030003	PACITAN	PRINGKUKU	SUGIHWARAS	70.02
29	3501030004	PACITAN	PRINGKUKU	JLUBANG	63.61
30	3501030005	PACITAN	PRINGKUKU	CANDI	67.80
31	3501030006	PACITAN	PRINGKUKU	POKO	68.05
32	3501030007	PACITAN	PRINGKUKU	DADAPAN	64.32
33	3501030008	PACITAN	PRINGKUKU	PRINGKUKU	63.51
34	3501030009	PACITAN	PRINGKUKU	SOBO	61.63
35	3501030010	PACITAN	PRINGKUKU	NGADIREJAN	73.73
36	3501030011	PACITAN	PRINGKUKU	GLINGGANGAN	57.47
37	3501030012	PACITAN	PRINGKUKU	PELEM	68.35
38	3501030013	PACITAN	PRINGKUKU	TAMANASRI	63.48
39	3501040003	PACITAN	PACITAN	KEMBANG	57.58
40	3501040004	PACITAN	PACITAN	SUKOHARJO	61.90
⋮	⋮	⋮	⋮	⋮	⋮
7721	3579030009	BATU	BUMIAJI	SUMBER BRANTAS	70.37

Lampiran 2. Data yang digunakan untuk membentuk model klasifikasi

No	Desa	Y	X1	X2	X3	X4	X5	...	X42
1	3501010001	0	0	0	1	1	1	...	1
2	3501010002	0	0	0	1	1	1	...	1
3	3501010003	0	0	0	0	0	1	...	1
4	3501010004	0	0	0	1	1	1	...	1
5	3501010005	0	0	0	0	1	1	...	1
6	3501010006	0	0	0	0	1	1	...	1
7	3501010007	0	0	0	1	1	1	...	1
8	3501010008	0	0	0	0	1	1	...	1
9	3501010009	0	0	0	1	1	1	...	1
10	3501010010	0	0	0	0	1	1	...	1
11	3501010011	0	0	0	0	0	1	...	0
12	3501010012	0	0	0	1	0	1	...	0
13	3501020001	0	0	0	0	0	1	...	1
14	3501020002	0	0	0	0	0	1	...	1
15	3501020003	0	0	0	0	1	1	...	1
16	3501020004	0	0	0	1	1	1	...	1
17	3501020005	0	0	0	0	0	1	...	1
18	3501020006	0	0	0	1	1	1	...	1
19	3501020007	0	0	0	1	1	1	...	1
20	3501020008	0	0	0	1	1	1	...	1
21	3501020009	0	0	0	1	1	1	...	1
22	3501020010	0	0	0	0	0	1	...	1
23	3501020011	0	0	0	0	1	1	...	1
24	3501020012	0	0	0	0	1	1	...	1
25	3501020013	0	0	0	1	1	1	...	1
26	3501030001	0	0	0	1	1	1	...	1
27	3501030002	0	0	0	0	1	1	...	1
28	3501030003	0	0	0	0	1	1	...	1
29	3501030004	0	0	0	1	1	1	...	1
30	3501030005	0	0	0	0	1	1	...	1
31	3501030006	0	0	0	1	1	1	...	0
32	3501030007	0	0	0	1	1	1	...	1
33	3501030008	0	0	0	1	1	1	...	1
34	3501030009	0	0	0	1	1	1	...	1
35	3501030010	0	0	0	0	0	1	...	1
36	3501030011	0	0	0	0	1	1	...	0
37	3501030012	0	0	0	1	1	1	...	1
38	3501030013	0	0	0	1	1	1	...	1
39	3501040003	0	0	0	1	1	1	...	1
40	3501040004	0	0	0	1	1	1	...	1
⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	...	⋮
7721	3579030009	0	0	0	0	1	1	...	0

Lampiran 3. Klasifikasi Desa Tertinggal dan Tidak Tertinggal menurut Kabupaten/Kota di Jawa Timur Tahun 2014

No	Kabupaten/Kota		Klasifikasi		Total
			Tidak Tertinggal	Tertinggal	
1.	Pacitan	Jumlah	166	0	166
		Persentase	100,0%	0,0%	100,0%
2.	Ponorogo	Jumlah	277	4	281
		Persentase	98,6%	1,4%	100,0%
3.	Trenggalek	Jumlah	146	6	152
		Persentase	96,1%	3,9%	100,0%
4.	Tulungagung	Jumlah	253	4	257
		Persentase	98,4%	1,6%	100,0%
5.	Blitar	Jumlah	219	1	220
		Persentase	99,5%	0,5%	100,0%
6.	Kediri	Jumlah	341	2	343
		Persentase	99,4%	0,6%	100,0%
7.	Malang	Jumlah	370	7	377
		Persentase	98,1%	1,9%	100,0%
8.	Lumajang	Jumlah	188	10	198
		Persentase	94,9%	5,1%	100,0%
9.	Jember	Jumlah	225	1	226
		Persentase	99,6%	0,4%	100,0%
10.	Banyuwangi	Jumlah	189	0	189
		Persentase	100,0%	0,0%	100,0%
11.	Bondowoso	Jumlah	192	17	209
		Persentase	91,9%	8,1%	100,0%
12.	Situbondo	Jumlah	120	12	132
		Persentase	90,9%	9,1%	100,0%
13.	Probolinggo	Jumlah	316	9	325
		Persentase	97,2%	2,8%	100,0%
14.	Pasuruan	Jumlah	329	12	341
		Persentase	96,5%	3,5%	100,0%
15.	Sidoarjo	Jumlah	321	1	322
		Persentase	99,7%	0,3%	100,0%
16.	Mojokerto	Jumlah	297	2	299
		Persentase	99,3%	0,7%	100,0%
17.	Jombang	Jumlah	301	1	302
		Persentase	99,7%	0,3%	100,0%
18.	Nganjuk	Jumlah	257	7	264
		Persentase	97,3%	2,7%	100,0%
19.	Madiun	Jumlah	196	2	198
		Persentase	99,0%	1,0%	100,0%

Lampiran 3. (Lanjutan)

No	Kabupaten/Kota		Klasifikasi		Total
			Tidak Tertinggal	Tertinggal	
20.	Magetan	Jumlah	207	0	207
		Persentase	100,0%	0,0%	100,0%
21.	Ngawi	Jumlah	211	2	213
		Persentase	99,1%	0,9%	100,0%
22.	Bojonegoro	Jumlah	415	4	419
		Persentase	99,0%	1,0%	100,0%
23.	Tuban	Jumlah	303	8	311
		Persentase	97,4%	2,6%	100,0%
24.	Lamongan	Jumlah	460	2	462
		Persentase	99,6%	0,4%	100,0%
25.	Gresik	Jumlah	330	0	330
		Persentase	100,0%	0,0%	100,0%
26.	Bangkalan	Jumlah	229	44	273
		Persentase	83,9%	16,1%	100,0%
27.	Sampang	Jumlah	166	14	180
		Persentase	92,2%	7,8%	100,0%
28.	Pamekasan	Jumlah	171	7	178
		Persentase	96,1%	3,9%	100,0%
29.	Sumenep	Jumlah	299	29	328
		Persentase	91,2%	8,8%	100,0%
30.	Batu	Jumlah	19	0	19
		Persentase	100,0%	0,0%	100,0%
Total		Jumlah	7513	208	7721
		Persentase	97,3%	2,7%	100,0%

Sumber : Bappenas

Keterangan: Kota di Jawa Timur yang mempunyai desa hanya Kota Batu. Selain Batu, semua kota di Jawa Timur terdiri dari kelurahan.

Lampiran 4. Jumlah dan Persentase Desa menurut Kategori Variabel Penelitian

Variabel	Kategori	Jumlah Desa	Persentase (%)
X1 Ketersediaan TK/RA/BA	Tersedia	7487	96,97
	Tidak tersedia	234	3,03
X2 Ketersediaan SD Sederajat	Tersedia	7681	99,48
	Tidak tersedia	40	0,52
X3 Ketersediaan SMP Sederajat	Tersedia	3928	50,87
	Tidak tersedia	3793	49,13
X4 Ketersediaan SMA Sederajat	Tersedia	2150	27,85
	Tidak tersedia	5571	72,15
X5 Ketersediaan Rumah Sakit	Tersedia	121	1,57
	Tidak tersedia	7600	98,43
X6 Ketersediaan Rumah Sakit Bersalin	Tersedia	237	3,07
	Tidak tersedia	7484	96,93
X7 Ketersediaan Puskesmas	Tersedia	2747	35,58
	Tidak tersedia	4974	64,42
X8 Ketersediaan Poliklinik/Balai Pengobatan	Tersedia	603	7,81
	Tidak tersedia	7118	92,19
X9 Ketersediaan Tempat Praktek Dokter	Tersedia	1694	21,94
	Tidak tersedia	6027	78,06
X10 Ketersediaan Tempat Praktek Bidan	Tersedia	6726	87,11
	Tidak tersedia	995	12,89
X11 Ketersediaan Poskesdes atau Polindes	Tersedia	6473	83,84
	Tidak tersedia	1248	16,16
X12 Ketersediaan Apotek	Tersedia	895	11,59
	Tidak tersedia	6826	88,41
X13 Ketersediaan Pertokoan, Minimarket atau Toko Kelontong	Tersedia	7636	98,90
	Tidak tersedia	85	1,10
X14 Ketersediaan Pasar	Tersedia	2234	28,93
	Tidak tersedia	5487	71,07
X15 Ketersediaan Restoran, Rumah Makan atau Warung Makan	Tersedia	7430	96,23
	Tidak tersedia	291	3,77
X16 Ketersediaan Akomodasi Hotel atau Penginapan	Tersedia	258	3,34
	Tidak tersedia	7463	96,66
X17 Ketersediaan Bank	Tersedia	1122	14,53
	Tidak tersedia	6599	85,47
X18 Elektrifikasi/ Persentase Keluarga Pengguna Listrik	> 80%	7708	99,83
	≤ 80%	13	0,17

Lampiran 4. (Lanjutan)

Variabel		Kategori	Jumlah Desa	Persentase (%)
X19	Kondisi Penerangan di Jalan Utama	Ada penerangan	7276	94,24
		Tidak ada penerangan	445	5,76
X20	Bahan Bakar Untuk Memasak	LPG atau gas kota	5566	72,09
		minyak tanah, kayu bakar, batu bara, dll	2155	27,91
X21	Sumber Air Untuk Minum	sumur, ledeng, atau air kemasan	6335	82,05
		air hujan, sungai/danau/kolam, atau mata air	1386	17,95
X22	Sumber Air Untuk Mandi/Cuci	sumur atau ledeng	6189	80,16
		air hujan, sungai/danau/kolam, atau mata air	1532	19,84
X23	Fasilitas Buang Air Besar	Jamban	6734	87,22
		Bukan jamban	987	12,78
X24	Kualitas Fasilitas Komunikasi Seluler	Ada sinyal	7672	99,37
		Tidak ada sinyal	49	0,63
X25	Ketersediaan Fasilitas Internet dan Pengiriman Pos atau Barang	Tersedia	922	11,94
		Tidak tersedia	6799	88,06
X26	Lalu Lintas dan Kualitas Jalan	Aspal/beton	6966	90,22
		Selain aspal/beton	755	9,78
X27	Aksesibilitas Jalan	Dapat dilalui	7577	98,13
		Tidak dapat dilalui	144	1,87
X28	Ketersediaan Angkutan Umum	Ada angkutan umum	4887	63,29
		Tidak ada angkutan umum	2834	36,71
X29	Operasional Angkutan Umum	Setiap hari	4224	54,71
		Tidak setiap hari	3497	45,29
X30	Waktu Tempuh per Kilometer Transportasi ke Kantor Camat	≤ 9 menit/km	1983	25,68
		> 9 menit/km	5738	74,32
X31	Biaya per Kilometer Transportasi ke Kantor Camat	≤ Rp. 7.000,-	7566	97,99
		> Rp. 7.000,-	155	2,01
X32	Waktu Tempuh per Kilometer Transportasi ke Kantor Bupati/Walikota	≤ 5 menit/km	5904	76,47
		> 5 menit/km	1817	23,53

Lampiran 4. (Lanjutan)

Variabel	Kategori	Jumlah Desa	Persentase (%)
X33 Biaya per Kilometer Transportasi ke Kantor Bupati/Walikota	≤ Rp. 3.000,-	7579	98,16
	> Rp. 3.000,-	142	1,84
X34 Penanganan Kejadian Luar Biasa (KLB)	Tidak ada KLB	6732	87,19
	Ada KLB	989	12,81
X35 Penanganan Gizi Buruk	Tidak ada penderita gizi buruk	5892	76,31
	Ada penderita gizi buruk	1829	23,69
X36 Ketersediaan Fasilitas Olah Raga	Tersedia	6954	90,07
	Tidak tersedia	767	9,93
X37 Keberadaan Kelompok Kegiatan Olah Raga	Ada	6535	84,64
	Tidak ada	1186	15,36
X38 Kelengkapan Pemerintahan Desa	Ada BPD dan kantor kepala desa	7250	93,90
	Tidak ada BPD atau kantor kepala desa	471	6,10
X39 Otonomi Desa	PAD ≥ 27%	2903	37,60
	PAD < 27%	4818	62,40
X40 Aset/Kekayaan Desa	Aset > 2 jenis	3877	50,21
	Aset ≤ 2 jenis	3844	49,79
X41 Kualitas SDM Kepala Desa	Pendidikan > SMA sederajat	2006	25,98
	Pendidikan ≤ SMA sederajat	5715	74,02
X42 Kualitas SDM Sekretaris Desa	Pendidikan > SMA sederajat	2097	27,16
	Pendidikan ≤ SMA sederajat	5624	72,84

Sumber: Hasil Pengolahan SPSS

Lampiran 5. Syntax R untuk klasifikasi dengan *TR-IRLS* dan *RE-WLR*

a. Syntax pembagian data

```
library(foreign)
data=read.spss('E:\\KULIAH/TESIS/Program/podes_r.sav', to.data.frame=TRUE)
splitDataFrame <- function(dataframe, seed = NULL, n = trainSize)
{
  if (!is.null(seed)) set.seed(seed)
  index <- 1:nrow(dataframe)
  trainindex <- sample(index, n)
  trainset <- dataframe[trainindex, ]
  testset <- dataframe[-trainindex, ]
  list(trainset = trainset, testset = testset)
}

dataList1 <- splitDataFrame(data, NULL, round(nrow(data) * 0.5))
train50 <- dataList1$trainset
test50 <- dataList1$testset
write.table(train50, 'E:\\KULIAH/TESIS/Program/podes_train50.txt', col.names=TRUE)
write.table(test50, 'E:\\KULIAH/TESIS/Program/podes_test50.txt', col.names=TRUE)

dataList2 <- splitDataFrame(data, NULL, round(nrow(data) * 0.6))
train60 <- dataList2$trainset
test40 <- dataList2$testset
write.table(train60, 'E:\\KULIAH/TESIS/Program/podes_train60.txt', col.names=TRUE)
write.table(test40, 'E:\\KULIAH/TESIS/Program/podes_test40.txt', col.names=TRUE)

dataList3 <- splitDataFrame(data, NULL, round(nrow(data) * 0.7))
train70 <- dataList3$trainset
test30 <- dataList3$testset
write.table(train70, 'E:\\KULIAH/TESIS/Program/podes_train70.txt', col.names=TRUE)
write.table(test30, 'E:\\KULIAH/TESIS/Program/podes_test30.txt', col.names=TRUE)

dataList4 <- splitDataFrame(data, NULL, round(nrow(data) * 0.8))
train80 <- dataList4$trainset
test20 <- dataList4$testset
write.table(train80, 'E:\\KULIAH/TESIS/Program/podes_train80.txt', col.names=TRUE)
write.table(test20, 'E:\\KULIAH/TESIS/Program/podes_test20.txt', col.names=TRUE)

dataList5 <- splitDataFrame(data, NULL, round(nrow(data) * 0.9))
train90 <- dataList5$trainset
test10 <- dataList5$testset
write.table(train90, 'E:\\KULIAH/TESIS/Program/podes_train90.txt', col.names=TRUE)
write.table(test10, 'E:\\KULIAH/TESIS/Program/podes_test10.txt', col.names=TRUE)

data50train = read.table('E:\\KULIAH/TESIS/Program/podes_train50.txt', header=TRUE)
data50test = read.table('E:\\KULIAH/TESIS/Program/podes_test50.txt', header=TRUE)
data60train = read.table('E:\\KULIAH/TESIS/Program/podes_train60.txt', header=TRUE)
data40test = read.table('E:\\KULIAH/TESIS/Program/podes_test40.txt', header=TRUE)
data70train = read.table('E:\\KULIAH/TESIS/Program/podes_train70.txt', header=TRUE)
data30test = read.table('E:\\KULIAH/TESIS/Program/podes_test30.txt', header=TRUE)
data80train = read.table('E:\\KULIAH/TESIS/Program/podes_train80.txt', header=TRUE)
data20test = read.table('E:\\KULIAH/TESIS/Program/podes_test20.txt', header=TRUE)
```

```
data90train = read.table('E:\\KULIAH/TESIS/Program/podes_train90.txt', header=TRUE)
data10test = read.table('E:\\KULIAH/TESIS/Program/podes_test10.txt', header=TRUE)
```

Pembagian Data Stratifikasi

```
library(foreign)
dataR1=read.spss('E:\\KULIAH/TESIS/Program/podes_r1.sav', to.data.frame=TRUE)
dataR0=read.spss('E:\\KULIAH/TESIS/Program/podes_r0.sav', to.data.frame=TRUE)
dataListP1 <- splitDataFrame(dataR1, NULL, round(nrow(dataR1) * 0.5))
dataListP2 <- splitDataFrame(dataR1, NULL, round(nrow(dataR1) * 0.6))
dataListP3 <- splitDataFrame(dataR1, NULL, round(nrow(dataR1) * 0.7))
dataListP4 <- splitDataFrame(dataR1, NULL, round(nrow(dataR1) * 0.8))
dataListP5 <- splitDataFrame(dataR1, NULL, round(nrow(dataR1) * 0.9))

dataListN1 <- splitDataFrame(dataR0, NULL, round(nrow(dataR0) * 0.5))
dataListN2 <- splitDataFrame(dataR0, NULL, round(nrow(dataR0) * 0.6))
dataListN3 <- splitDataFrame(dataR0, NULL, round(nrow(dataR0) * 0.7))
dataListN4 <- splitDataFrame(dataR0, NULL, round(nrow(dataR0) * 0.8))
dataListN5 <- splitDataFrame(dataR0, NULL, round(nrow(dataR0) * 0.9))

trainP50 <- dataListP1$trainset
testP50 <- dataListP1$testset
trainP60 <- dataListP2$trainset
testP40 <- dataListP2$testset
trainP70 <- dataListP3$trainset
testP30 <- dataListP3$testset
trainP80 <- dataListP4$trainset
testP20 <- dataListP4$testset
trainP90 <- dataListP5$trainset
testP10 <- dataListP5$testset

trainN50 <- dataListN1$trainset
testN50 <- dataListN1$testset
trainN60 <- dataListN2$trainset
testN40 <- dataListN2$testset
trainN70 <- dataListN3$trainset
testN30 <- dataListN3$testset
trainN80 <- dataListN4$trainset
testN20 <- dataListN4$testset
trainN90 <- dataListN5$trainset
testN10 <- dataListN5$testset

trainS50 = rbind(trainP50, trainN50)
testS50 = rbind(testP50, testN50)
trainS60 = rbind(trainP60, trainN60)
testS40 = rbind(testP40, testN40)
trainS70 = rbind(trainP70, trainN70)
testS30 = rbind(testP30, testN30)
trainS80 = rbind(trainP80, trainN80)

testS20 = rbind(testP20, testN20)
trainS90 = rbind(trainP90, trainN90)
```

```
testS10 = rbind(testP10, testN10)
```

```
write.table(trainS50, 'E:\\KULIAH\\TESIS\\Program\\podes_trainS50.txt', col.names=TRUE)
write.table(testS50, 'E:\\KULIAH\\TESIS\\Program\\podes_testS50.txt', col.names=TRUE)
write.table(trainS60, 'E:\\KULIAH\\TESIS\\Program\\podes_trainS60.txt', col.names=TRUE)
write.table(testS40, 'E:\\KULIAH\\TESIS\\Program\\podes_testS40.txt', col.names=TRUE)
write.table(trainS70, 'E:\\KULIAH\\TESIS\\Program\\podes_trainS70.txt', col.names=TRUE)
write.table(testS30, 'E:\\KULIAH\\TESIS\\Program\\podes_testS30.txt', col.names=TRUE)
write.table(trainS80, 'E:\\KULIAH\\TESIS\\Program\\podes_trainS80.txt', col.names=TRUE)
write.table(testS20, 'E:\\KULIAH\\TESIS\\Program\\podes_testS20.txt', col.names=TRUE)
write.table(trainS90, 'E:\\KULIAH\\TESIS\\Program\\podes_trainS90.txt', col.names=TRUE)
write.table(testS10, 'E:\\KULIAH\\TESIS\\Program\\podes_testS10.txt', col.names=TRUE)
```

```
dataS50train = read.table('E:\\KULIAH\\TESIS\\Program\\podes_trainS50.txt', header=TRUE)
dataS50test = read.table('E:\\KULIAH\\TESIS\\Program\\podes_testS50.txt', header=TRUE)
dataS60train = read.table('E:\\KULIAH\\TESIS\\Program\\podes_trainS60.txt', header=TRUE)
dataS40test = read.table('E:\\KULIAH\\TESIS\\Program\\podes_testS40.txt', header=TRUE)
dataS70train = read.table('E:\\KULIAH\\TESIS\\Program\\podes_trainS70.txt', header=TRUE)
dataS30test = read.table('E:\\KULIAH\\TESIS\\Program\\podes_testS30.txt', header=TRUE)
dataS80train = read.table('E:\\KULIAH\\TESIS\\Program\\podes_trainS80.txt', header=TRUE)
dataS20test = read.table('E:\\KULIAH\\TESIS\\Program\\podes_testS20.txt', header=TRUE)
dataS90train = read.table('E:\\KULIAH\\TESIS\\Program\\podes_trainS90.txt', header=TRUE)
dataS10test = read.table('E:\\KULIAH\\TESIS\\Program\\podes_testS10.txt', header=TRUE)
```

b. Syntax program klasifikasi *TR-IRLS* (Contoh untuk partisi data 50:50)

```
{
for(lambda in 1:10){
  ptm <- proc.time()

  data1 = data50train
  data2 = data50test
  tr = 50
  maxIRLS = 30
  maxCG = 200

  mf <- model.frame(Y ~ X1 + X2 + X3 + X4 + X5 + X6 + X7 + X8 + X9 + X10 + X11 + X12 +
X13 + X14 + X15 + X16 + X17 + X18 + X19 + X20 + X21 + X22 + X23 + X24 + X25 + X26 + X27 + X28 +
X29 + X30 + X31 + X32 + X33 + X34 + X35 + X36 + X37 + X38 + X39 + X40 + X41 + X42,data=data1)
  y <- model.response(mf)
  X <- model.matrix(Y ~ X1 + X2 + X3 + X4 + X5 + X6 + X7 + X8 + X9 + X10 + X11 + X12 + X13
+ X14 + X15 + X16 + X17 + X18 + X19 + X20 + X21 + X22 + X23 + X24 + X25 + X26 + X27 + X28 + X29
+ X30 + X31 + X32 + X33 + X34 + X35 + X36 + X37 + X38 + X39 + X40 + X41 + X42,data=data1)

  mft <- model.frame(Y ~ X1 + X2 + X3 + X4 + X5 + X6 + X7 + X8 + X9 + X10 + X11 + X12 +
X13 + X14 + X15 + X16 + X17 + X18 + X19 + X20 + X21 + X22 + X23 + X24 + X25 + X26 + X27 + X28 +
X29 + X30 + X31 + X32 + X33 + X34 + X35 + X36 + X37 + X38 + X39 + X40 + X41 + X42,data=data2)
  yt <- model.response(mft)
  Xt <- model.matrix(Y ~ X1 + X2 + X3 + X4 + X5 + X6 + X7 + X8 + X9 + X10 + X11 + X12 +
X13 + X14 + X15 + X16 + X17 + X18 + X19 + X20 + X21 + X22 + X23 + X24 + X25 + X26 + X27 + X28 +
X29 + X30 + X31 + X32 + X33 + X34 + X35 + X36 + X37 + X38 + X39 + X40 + X41 + X42,data=data2)
```

```

beta = rep(0, ncol(X))
Beta = matrix(beta)
deltadev = 1
c = 0
I = diag(ncol(X))
I[1,1]=0
ts=100-tr
cat("Hasil model klasifikasi TR-IRLS data partisi ",tr,":",ts, " dengan lambda ", lambda,
"sbb:\n")

while (abs(deltadev) > 0.01 & c <= maxIRLS){

  p = matrix(nrow=nrow(X), ncol=1)
  V = matrix(data=0, nrow=nrow(X), ncol=nrow(X))
  Z = matrix(nrow=nrow(X), ncol=1)
  for (i in 1:nrow(X)){
    p[i]=(exp (X[i,] %**% Beta))/(1 + exp (X[i,] %**% Beta))
    V[i,i] = p[i] * (1-p[i])
    Z[i] = X[i,] %**% Beta + (y[i] - p[i])/V[i,i]
  }

  if (c == 0){
    logB = 0
    for (i in 1:nrow(X)){
      logB1 = (y[i] * log(p[i]) + (1 - y[i]) * log(1-p[i])) + logB
    }
    logB = logB1
    dev = (-2) * (logB - lambda/2 * sum(Beta^2))
  }

  A = (t(X) %**% V %**% X) + lambda * I
  b = t(X) %**% V %**% Z

  r = b - A %**% Beta
  t = 0
  sumr2 = sum(r^2)
  d = r

  while (sumr2 > 0.005 & t <= maxCG){

    Ad = A %**% d
    s = as.numeric((t(r) %**% r)/(t(d) %**% Ad))
    Beta = Beta + s * d
    r1 = r - s * Ad
    a = as.numeric((t(r1) %**% r1)/(t(r) %**% r))
    d1 = r1 + a * d
    sumr2 = sum(r1^2)
    t=t+1
    r = r1
    d = d1

  }
}

```



```

if (t > maxCG) print("Belum konvergen, tambah nilai max iterasi CG.")

logB = 0
for (i in 1:nrow(X)){
  p[i]=(exp (X[i,] %*% Beta))/(1 + exp (X[i,] %*% Beta))
  logB1 = (y[i] * log(p[i]) + (1 - y[i]) * log(1-p[i])) + logB
  logB = logB1
}

devold = dev
dev = (-2) * (logB - lambda/2 * sum(Beta^2))
deltadev = (devold - dev)/dev
c = c + 1
cat("Iterasi ",c," , dev = ",dev," , deltadev = ",deltadev,"\n")
}
if (c > maxIRLS) print("Belum konvergen, tambah nilai max iterasi IRLS.")

phat = matrix(nrow=nrow(X), ncol=1)
yhat = matrix(nrow=nrow(X), ncol=1)
fitt = matrix(nrow=nrow(X), ncol=1)

for (k in 1:nrow(X)){
  phat[k]=(exp (X[k,] %*% Beta))/(1 + exp (X[k,] %*% Beta))
  if (phat[k] < 0.5)yhat[k]=0
  else yhat[k] = 1
  if (yhat[k] == y[k]) fitt[k] = 1
  else fitt[k] = 0
}
tepat1 = mean(fitt)

pthat = matrix(nrow=nrow(Xt), ncol=1)
ythat = matrix(nrow=nrow(Xt), ncol=1)
fit = matrix(nrow=nrow(Xt), ncol=1)

for (k in 1:nrow(Xt)){
  pthat[k]=(exp (Xt[k,] %*% Beta))/(1 + exp (Xt[k,] %*% Beta))
  if (pthat[k] < 0.5)ythat[k]=0
  else ythat[k] = 1
  if (ythat[k] == yt[k]) fit[k] = 1
  else fit[k] = 0
}
tepat = mean(fit)

cat("Hasil estimasi parameter Beta:\n")
print(Beta)
cat("Jumlah iterasi = ",c,"\n")
cat("Confusion matrix Data Training:\n")
library(crossval)
cm1 = confusionMatrix(y, yhat, negative=0)
print(cm1)
de1 = diagnosticErrors(cm1)
print(de1)

```

```

        cat("Ketepatan klasifikasi TR-IRLS Data Training (" ,tr,":",ts, ") dan lambda ", lambda,
"yaitu ", tepat1,"\n")

        library(verification)
        rocArea1= roc.area(y, yhat)
        cat("ROC Area Data Training sbb:\n")
        print(rocArea1)

        cat("Confusion matrix Data Testing:\n")
        cm = confusionMatrix(yt, ythat, negative=0)
        print(cm)
        de = diagnosticErrors(cm)
        print(de)

        cat("Ketepatan klasifikasi TR-IRLS Data Testing (" ,tr,":",ts, ") dan lambda ", lambda,
"yaitu ", tepat, "\n")

        rocArea= roc.area(yt, ythat)
        cat("ROC Area Data Testing sbb:\n")
        print(rocArea)

        etm = proc.time() - ptm
        cat("\n")
        print(etm)
        cat("-----\n")
        lambda = lambda + 1
    }
}

```

c. Syntax program klasifikasi *RE-WLR* (Contoh untuk partisi data 50:50)

```

{
for(lambda in 1:10){
    ptm <- proc.time()

    data1 = data50train
    data2 = data50test
    tr = 50
    maxIRLS = 30
    maxCG1 = 200
    maxCG2 = 200

    mf <- model.frame(Y ~ X1 + X2 + X3 + X4 + X5 + X6 + X7 + X8 + X9 + X10 + X11 + X12 +
X13 + X14 + X15 + X16 + X17 + X18 + X19 + X20 + X21 + X22 + X23 + X24 + X25 + X26 + X27 + X28 +
X29 + X30 + X31 + X32 + X33 + X34 + X35 + X36 + X37 + X38 + X39 + X40 + X41 + X42,data=data1)
    y <- model.response(mf)
    X <- model.matrix(Y ~ X1 + X2 + X3 + X4 + X5 + X6 + X7 + X8 + X9 + X10 + X11 + X12 + X13
+ X14 + X15 + X16 + X17 + X18 + X19 + X20 + X21 + X22 + X23 + X24 + X25 + X26 + X27 + X28 + X29
+ X30 + X31 + X32 + X33 + X34 + X35 + X36 + X37 + X38 + X39 + X40 + X41 + X42,data=data1)

```

```

mft <- model.frame(Y ~ X1 + X2 + X3 + X4 + X5 + X6 + X7 + X8 + X9 + X10 + X11 + X12 +
X13 + X14 + X15 + X16 + X17 + X18 + X19 + X20 + X21 + X22 + X23 + X24 + X25 + X26 + X27 + X28 +
X29 + X30 + X31 + X32 + X33 + X34 + X35 + X36 + X37 + X38 + X39 + X40 + X41 + X42,data=data2)
yt <- model.response(mft)
Xt <- model.matrix(Y ~ X1 + X2 + X3 + X4 + X5 + X6 + X7 + X8 + X9 + X10 + X11 + X12 +
X13 + X14 + X15 + X16 + X17 + X18 + X19 + X20 + X21 + X22 + X23 + X24 + X25 + X26 + X27 + X28 +
X29 + X30 + X31 + X32 + X33 + X34 + X35 + X36 + X37 + X38 + X39 + X40 + X41 + X42,data=data2)

beta = rep(0, ncol(X))
Beta = matrix(beta)
deltadev = 1
c = 0
I = diag(ncol(X))
I[1,1]=0
tau = 0.0269
ybar = mean(y)
w1 = tau/ybar
w0 = (1 - tau)/(1 - ybar)
ts=100-tr
cat("Hasil model klasifikasi RE-WLR data partisi ",tr,":",ts, " dengan lambda ", lambda,
"sbb:\n")

while (abs(deltadev) > 0.01 & c <= maxIRLS){

  if (c > 0)Beta = uBeta

  p = matrix(nrow=nrow(X), ncol=1)
  V = matrix(data=0, nrow=nrow(X), ncol=nrow(X))
  W = matrix(data=0, nrow=nrow(X), ncol=nrow(X))
  Z = matrix(nrow=nrow(X), ncol=1)
  D = matrix(data=0, nrow=nrow(X), ncol=nrow(X))
  for (i in 1:nrow(X)){
    p[i] = (exp (X[i,] %*% Beta))/(1 + exp (X[i,] %*% Beta))
    V[i,i] = p[i] * (1-p[i])
    W[i,i] = (w1 * y[i]) + (w0 * (1 - y[i]))
    Z[i] = X[i,] %*% Beta + (y[i] - p[i])/V[i,i]
    D[i,i] = V[i,i] * W[i,i]
  }

  if (c == 0){
    logB = 0
    for (i in 1:nrow(X)){
      logB1 = W[i,i]*(y[i] * log(p[i]) + (1 - y[i]) * log(1-p[i])) + logB
    }
    logB = logB1
    dev = (-2) * (logB - lambda/2 * sum(Beta^2))
  }

  Q = X %*% solve((t(X) %*% D %*% X) + lambda * I) %*% t(X)
  q = diag(Q)

  E = matrix(nrow=nrow(X), ncol=1)
  for (i in 1:nrow(X)){

```

```

        E[i] = 0.5 * q[i] * ((1+w1) * p[i] - w1)
    }

A = (t(X) %*% D %*% X) + lambda * I
b = t(X) %*% D %*% Z
C = t(X) %*% D %*% E

r = b - A %*% Beta
t = 0
sumr2 = sum(r^2)
d = r

while (sumr2 > 0.005 & t <= maxCG1){
    Ad = A %*% d
    s = as.numeric((t(r) %*% r)/(t(d) %*% Ad))
    Beta = Beta + s * d
    r1 = r - s * Ad
    a = as.numeric((t(r1) %*% r1)/(t(r) %*% r))
    d1 = r1 + a * d
    sumr2 = sum(r1^2)
    t = t + 1
    r = r1
    d = d1
}
if (t > maxCG1) print("Belum konvergen, tambah nilai max iterasi CG1.")

bbeta = rep(0, length(C))
BBeta = matrix (bbeta)
rb = C - A %*% BBeta
h = 0
sumrb2 = sum(rb^2)
db = rb

while (sumrb2 > 0.005 & h <= maxCG2){
    Adb = A %*% db
    sb = as.numeric((t(rb) %*% rb)/(t(db) %*% Adb))
    BBeta = BBeta + sb * db
    rb1 = rb - sb * Adb
    ab = as.numeric((t(rb1) %*% rb1)/(t(rb) %*% rb))
    db1 = rb1 + ab * db
    sumrb2 = sum(rb1^2)
    h = h + 1
    rb = rb1
    db = db1
}
if (h > maxCG2) print("Belum konvergen, tambah nilai max iterasi CG2.")

uBeta = Beta - BBeta
logB = 0
for (i in 1:nrow(X)){
    p[i]=(exp (X[i,] %*% uBeta))/(1 + exp (X[i,] %*% uBeta))
    logB1 = W[i,i]*(y[i] * log(p[i]) + (1 - y[i]) * log(1-p[i])) + logB
    logB = logB1
}

```

```

    }

    devold = dev
    dev = (-2) * (logB - lambda/2 * sum(uBeta^2))
    deltadev = (dev - devold)/dev
    c = c + 1
    cat("Iterasi ",c," , dev = ",dev," , deltadev = ",deltadev,"\n")
}
if (c > maxIRLS) print("Belum konvergen, tambah nilai max iterasi IRLS.")

phat = matrix(nrow=nrow(X), ncol=1)
yhat = matrix(nrow=nrow(X), ncol=1)
fitt = matrix(nrow=nrow(X), ncol=1)

for (k in 1:nrow(X)){
  phat[k]=(exp (X[k,] %*% uBeta))/(1 + exp (X[k,] %*% uBeta))
  if (phat[k] < 0.5)yhat[k]=0
  else yhat[k] = 1
  if (yhat[k] == y[k]) fitt[k] = 1
  else fitt[k] = 0
}
tepat1 = mean(fitt)

pthat = matrix(nrow=nrow(Xt), ncol=1)
ythat = matrix(nrow=nrow(Xt), ncol=1)
fit = matrix(nrow=nrow(Xt), ncol=1)

for (k in 1:nrow(Xt)){
  pthat[k]=(exp (Xt[k,] %*% uBeta))/(1 + exp (Xt[k,] %*% uBeta))
  if (pthat[k] < 0.5)ythat[k]=0
  else ythat[k] = 1
  if (ythat[k] == yt[k]) fit[k] = 1
  else fit[k] = 0
}
tepat = mean(fit)

cat("Hasil estimasi parameter Beta:\n")
print(uBeta)
cat("Jumlah iterasi = ",c,"\n")
cat("Confusion matrix Data Training:\n")
library(crossval)
cm1 = confusionMatrix(y, yhat, negative=0)
print(cm1)
de1 = diagnosticErrors(cm1)
print(de1)

cat("Ketepatan klasifikasi RE-WLR Data Training (" ,tr,":",ts, ") dan lambda ", lambda,
"yaitu ", tepat1,"\n")

library(verification)
rocArea1= roc.area(y, yhat)
cat("ROC Area Data Training sbb:\n")

```

```

print(rocArea1)

cat("Confusion matrix Data Testing:\n")
cm = confusionMatrix(yt, ythat, negative=0)
print(cm)
de = diagnosticErrors(cm)
print(de)

cat("Ketepatan klasifikasi RE-WLR Data Testing (",tr,".",ts, ") dan lambda ", lambda,
"yaitu ", tepat,"\n")

rocArea= roc.area(yt, ythat)
cat("ROC Area Data Testing sbb:\n")
print(rocArea)

etm = proc.time() - ptm
cat("\n")
print(etm)
cat("-----\n")
lambda = lambda + 1
}
}

```

Lampiran 6. Contoh output dari klasifikasi dengan *TR-IRLS* dan *RE-WLR* menggunakan R

a. Output *TR-IRLS*

Hasil model klasifikasi TR-IRLS data partisi 50 : 50 dengan lambda 1 sbb:

Iterasi 1 , dev = 1337.835 , deltadev = 2.999817
Iterasi 2 , dev = 747.7728 , deltadev = 0.7890933
Iterasi 3 , dev = 542.5619 , deltadev = 0.3782258
Iterasi 4 , dev = 475.5763 , deltadev = 0.1408515
Iterasi 5 , dev = 490.9808 , deltadev = -0.03137503
Iterasi 6 , dev = 560.5486 , deltadev = -0.1241066
Iterasi 7 , dev = 619.3391 , deltadev = -0.09492451
Iterasi 8 , dev = 628.8861 , deltadev = -0.01518078
Iterasi 9 , dev = 628.8861 , deltadev = 0

Hasil estimasi parameter Beta:

[,1]

(Intercept) -15.834832600

X1	1.692678269
X2	0.122924983
X3	0.302001733
X4	1.208228396
X5	-0.004167312
X6	0.022789191
X7	2.712649886
X8	0.583866417
X9	0.717517289
X10	0.653545676
X11	-0.106667749
X12	0.101871226
X13	-0.789764673
X14	0.249159074
X15	1.034507640
X16	0.312213105
X17	0.376772900
X18	0.221018935
X19	1.653289221
X20	1.678925701
X21	1.253760977
X22	-0.149668983
X23	1.084129994
X24	1.148573944
X25	0.395231367
X26	0.204514765
X27	1.369649632
X28	0.567822280
X29	0.687333045
X30	-0.993527420
X31	1.404158408
X32	-0.367491413
X33	0.090513452
X34	0.106571911
X35	-0.186146867

```

X36      0.552315359
X37      0.576926139
X38      1.558027033
X39      1.310188359
X40      0.743951635
X41      0.286717856
X42      0.999717711
Jumlah iterasi = 9
Confusion matrix Data Training:
  FP  TP  TN  FN
  16  49 3743  52
attr(,"negative")
[1] 0
      acc  sens  spec  ppv  npv  lor
0.9823834 0.4851485 0.9957435 0.7538462 0.9862978 5.3956306
attr(,"negative")
[1] 0
Ketepatan klasifikasi TR-IRLS Data Training ( 50 : 50 ) dan lambda 1 yaitu 0.9823834
ROC Area Data Training sbb:
$A
[1] 0.740446

$n.total
[1] 3860

$n.events
[1] 101

$n.noevents
[1] 3759

$p.value
[1] 5.976026e-301

Confusion matrix Data Testing:
  FP  TP  TN  FN
  21  43 3733  64
attr(,"negative")
[1] 0
      acc  sens  spec  ppv  npv  lor
0.9779850 0.4018692 0.9944060 0.6718750 0.9831446 4.7827621
attr(,"negative")
[1] 0
Ketepatan klasifikasi TR-IRLS Data Testing ( 50 : 50 ) dan lambda 1 yaitu 0.977985
ROC Area Data Testing sbb:
$A
[1] 0.6981376

$n.total
[1] 3861

$n.events
[1] 107

```


\$n.noevents

[1] 3754

\$p.value

[1] 3.435158e-220

user system elapsed

19.16 0.45 19.89

b. Output *RE-WLR*

Hasil model klasifikasi RE-WLR data partisi 50 : 50 dengan lambda 1 sbb:

Iterasi 1 , dev = 1345.209 , deltadev = -2.977892

Iterasi 2 , dev = 763.6789 , deltadev = -0.7614852

Iterasi 3 , dev = 559.0396 , deltadev = -0.366055

Iterasi 4 , dev = 483.5962 , deltadev = -0.156005

Iterasi 5 , dev = 472.8241 , deltadev = -0.02278246

Iterasi 6 , dev = 497.8359 , deltadev = 0.05024112

Iterasi 7 , dev = 536.052 , deltadev = 0.07129178

Iterasi 8 , dev = 567.6942 , deltadev = 0.05573796

Iterasi 9 , dev = 582.5567 , deltadev = 0.02551256

Iterasi 10 , dev = 587.0599 , deltadev = 0.007670767

Hasil estimasi parameter Beta:

[,1]

(Intercept) -14.173410467

X1 1.605793404

X2 0.195287548

X3 0.276661408

X4 1.125656743

X5 -0.015739455

X6 -0.011667895

X7 2.538984217

X8 0.484318279

X9 0.571145752

X10 0.618981613

X11 -0.061690232

X12 -0.005038648

X13 -0.754170798

X14 0.197138678

X15 0.984046710

X16 0.228393955

X17 0.229700964

X18 0.239878746

X19 1.580594592

X20 1.582412249

X21 1.204022044

X22 -0.161427647

X23 1.027754585

X24 1.138747526

X25 0.251628110

X26 0.199952566

```

X27      1.356161035
X28      0.510116845
X29      0.664578011
X30     -0.938510974
X31      1.406562744
X32     -0.314787809
X33      0.084260157
X34      0.137688730
X35     -0.167236468
X36      0.510126795
X37      0.555552239
X38      1.450156578
X39      1.242185961
X40      0.680587059
X41      0.239290913
X42      0.950633999

```

Jumlah iterasi = 10

Confusion matrix Data Training:

```

FP TP TN FN
16 49 3743 52

```

attr("negative")

[1] 0

```

acc sens spec ppv npv lor
0.9823834 0.4851485 0.9957435 0.7538462 0.9862978 5.3956306

```

attr("negative")

[1] 0

Ketepatan klasifikasi RE-WLR Data Training (50 : 50) dan lambda 1 yaitu 0.9823834

ROC Area Data Training sbb:

\$A

[1] 0.740446

\$n.total

[1] 3860

\$n.events

[1] 101

\$n.noevents

[1] 3759

\$p.value

[1] 5.976026e-301

Confusion matrix Data Testing:

```

FP TP TN FN
23 46 3731 61

```

attr("negative")

[1] 0

```

acc sens spec ppv npv lor
0.9782440 0.4299065 0.9938732 0.6666667 0.9839135 4.8067049

```

attr("negative")

[1] 0

Ketepatan klasifikasi RE-WLR Data Testing (50 : 50) dan lambda 1 yaitu 0.978244

ROC Area Data Testing sbb:

\$A

[1] 0.7118899

\$n.total

[1] 3861

\$n.events

[1] 107

\$n.noevents

[1] 3754

\$p.value

[1] 9.975104e-234

user system elapsed

51.86 1.36 53.46

c. Output *RE-WLR* partisi data random dengan ukuran ketepatan klasifikasi optimal

Hasil model klasifikasi RE-WLR data partisi 80 : 20 dengan lambda 1 sbb:

Iterasi 1 , dev = 2158.565 , deltadev = -2.967052

Iterasi 2 , dev = 1227.997 , deltadev = -0.7577936

Iterasi 3 , dev = 898.9564 , deltadev = -0.3660251

Iterasi 4 , dev = 771.3016 , deltadev = -0.1655057

Iterasi 5 , dev = 745.1788 , deltadev = -0.03505576

Iterasi 6 , dev = 776.0681 , deltadev = 0.03980231

Iterasi 7 , dev = 826.4121 , deltadev = 0.06091882

Iterasi 8 , dev = 866.2398 , deltadev = 0.04597767

Iterasi 9 , dev = 882.4265 , deltadev = 0.01834346

Iterasi 10 , dev = 886.0285 , deltadev = 0.004065239

Hasil estimasi parameter Beta:

[,1]

(Intercept) -15.94903222

X1 1.18081160

X2 -0.10711376

X3 0.64811814

X4 1.26837481

X5 0.12994593

X6 0.14208049

X7 2.89152787

X8 0.60468370

X9 0.84983308

X10 0.50319082

X11 0.22080506

X12 0.04016721

X13 -0.07174735

X14 0.07103133

X15 0.72525118

X16 -0.08218794

```

X17      0.28964751
X18      0.34089745
X19      1.34128968
X20      1.71287467
X21      1.19708282
X22      0.10166546
X23      0.59630616
X24      1.36875489
X25      1.02636249
X26      0.80920328
X27      1.37022066
X28      1.05706836
X29      0.28958918
X30     -1.02434392
X31      1.28317778
X32     -0.72527261
X33     -0.00258227
X34      0.17774125
X35      0.03262516
X36      0.62332951
X37      0.44817788
X38      1.82856650
X39      1.08658416
X40      0.36744685
X41      0.15422046
X42      1.16240207

```

Jumlah iterasi = 10

Confusion matrix Data Training:

```

FP TP TN FN
22 78 5987 90

```

```

attr(,"negative")

```

```

[1] 0

```

```

      acc  sens  spec  ppv  npv  lor
0.9818682 0.4642857 0.9963388 0.7800000 0.9851901 5.4632024

```

```

attr(,"negative")

```

```

[1] 0

```

Ketepatan klasifikasi RE-WLR Data Training (80 : 20) dan lambda 1 yaitu 0.9818682

ROC Area Data Training sbb:

```

$A

```

```

[1] 0.7303123

```

```

$n.total

```

```

[1] 6177

```

```

$n.events

```

```

[1] 168

```

```

$n.noevents

```

```

[1] 6009

```

```

$p.value

```

```

[1] 0

```

Confusion matrix Data Testing:

```
FP TP TN FN
10 18 1494 22
```

```
attr("negative")
```

```
[1] 0
```

```
acc sens spec ppv npv lor
```

```
0.9792746 0.4500000 0.9933511 0.6428571 0.9854881 4.8059566
```

```
attr("negative")
```

```
[1] 0
```

Ketepatan klasifikasi RE-WLR Data Testing (80 : 20) dan lambda 1 yaitu 0.9792746

ROC Area Data Testing sbb:

```
$A
```

```
[1] 0.7216755
```

```
$n.total
```

```
[1] 1544
```

```
$n.events
```

```
[1] 40
```

```
$n.noevents
```

```
[1] 1504
```

```
$p.value
```

```
[1] 8.924344e-96
```

```
user system elapsed
```

```
130.61 2.92 134.55
```

d. Output *RE-WLR* pada partisi data stratifikasi dengan ukuran ketepatan klasifikasi optimal

Hasil model klasifikasi RE-WLR data partisi 50 : 50 dengan lambda 1 sbb:

Iterasi 1 , dev = 1344.105 , deltadev = -2.981159

Iterasi 2 , dev = 767.0894 , deltadev = -0.7522143

Iterasi 3 , dev = 568.5082 , deltadev = -0.3493024

Iterasi 4 , dev = 496.7765 , deltadev = -0.1443943

Iterasi 5 , dev = 487.9953 , deltadev = -0.01799432

Iterasi 6 , dev = 514.0066 , deltadev = 0.05060487

Iterasi 7 , dev = 551.6043 , deltadev = 0.0681607

Iterasi 8 , dev = 581.9933 , deltadev = 0.05221532

Iterasi 9 , dev = 596.8158 , deltadev = 0.02483609

Iterasi 10 , dev = 601.3801 , deltadev = 0.007589643

Hasil estimasi parameter Beta:

```
[,1]
```

```
(Intercept) -14.105355821
```

```
X1 1.002348205
```

```
X2 -0.030091898
```

```
X3 0.619762328
```

```
X4 1.373525816
```

```
X5 -0.026307559
```

```

X6      -0.013371510
X7      2.132804073
X8      0.453778094
X9      1.080345654
X10     0.445261260
X11     -0.287133636
X12     0.007103323
X13     -0.114044401
X14     -0.083445958
X15     0.636510304
X16     -0.439340846
X17     0.175352636
X18     -0.077283837
X19     1.185860637
X20     1.558205202
X21     0.974970883
X22     -0.083171652
X23     0.905328071
X24     0.876978737
X25     0.909744517
X26     0.686871142
X27     1.265495259
X28     0.840397572
X29     0.377412832
X30     -0.792570973
X31     0.918756494
X32     -0.563982866
X33     0.382473113
X34     -0.115308930
X35     0.059408109
X36     0.868091685
X37     0.781916407
X38     1.446311121
X39     0.705186075
X40     0.427070186
X41     0.343280310
X42     1.271462543

```

Jumlah iterasi = 10

Confusion matrix Data Training:

```

FP TP TN FN
14 48 3742 56

```

attr(,"negative")

[1] 0

```

acc sens spec ppv npv lor
0.9818653 0.4615385 0.9962726 0.7741935 0.9852554 5.4341675

```

attr(,"negative")

[1] 0

Ketepatan klasifikasi RE-WLR Data Training (50 : 50) dan lambda 1 yaitu 0.9818653

ROC Area Data Training sbb:

\$A

[1] 0.7289055

\$n.total

[1] 3860

\$n.events

[1] 104

\$n.noevents

[1] 3756

\$p.value

[1] 4.905818e-294

Confusion matrix Data Testing:

FP TP TN FN

15 47 3742 57

attr("negative")

[1] 0

acc sens spec ppv npv lor

0.9813520 0.4519231 0.9960075 0.7580645 0.9849961 5.3264216

attr("negative")

[1] 0

Ketepatan klasifikasi RE-WLR Data Testing (50 : 50) dan lambda 1 yaitu 0.981352

ROC Area Data Testing sbb:

\$A

[1] 0.7239653

\$n.total

[1] 3861

\$n.events

[1] 104

\$n.noevents

[1] 3757

\$p.value

[1] 1.167211e-281

user system elapsed

54.29 2.18 57.91

Lampiran 7. Confusion Matrix Model *TR-IRLS* dan *RE-WLR* Data Training

λ	Partisi Data	Kelas Sebenarnya	Kelas Prediksi <i>TR-IRLS</i>		Kelas Prediksi <i>RE-WLR</i>	
			Tertinggal	Tidak tertinggal	Tertinggal	Tidak tertinggal
1	50:50	Tertinggal	49	52	49	52
		Tidak tertinggal	16	3743	16	3743
	60:40	Tertinggal	62	80	61	81
		Tidak tertinggal	17	4474	18	4473
	70:30	Tertinggal	69	84	71	82
		Tidak tertinggal	21	5231	20	5232
	80:20	Tertinggal	69	99	78	90
		Tidak tertinggal	18	5991	22	5987
2	50:50	Tertinggal	87	101	87	101
		Tidak tertinggal	26	6735	26	6735
	50:50	Tertinggal	44	57	46	55
		Tidak tertinggal	11	3748	15	3744
	60:40	Tertinggal	60	82	55	87
		Tidak tertinggal	17	4474	14	4477
	70:30	Tertinggal	66	87	65	88
		Tidak tertinggal	21	5231	19	5233
3	80:20	Tertinggal	69	99	69	99
		Tidak tertinggal	18	5991	19	5990
	90:10	Tertinggal	77	111	73	115
		Tidak tertinggal	21	6740	20	6741
	50:50	Tertinggal	30	71	39	62
		Tidak tertinggal	6	3753	9	3750
	60:40	Tertinggal	54	88	48	94
		Tidak tertinggal	16	4475	13	4478
4	70:30	Tertinggal	63	90	59	94
		Tidak tertinggal	17	5235	17	5235
	80:20	Tertinggal	49	52	49	52
		Tidak tertinggal	16	3743	16	3743
	90:10	Tertinggal	62	80	61	81
		Tidak tertinggal	17	4474	18	4473
	50:50	Tertinggal	69	84	71	82
		Tidak tertinggal	21	5231	20	5232
4	60:40	Tertinggal	69	99	78	90
		Tidak tertinggal	18	5991	22	5987
	70:30	Tertinggal	87	101	87	101
		Tidak tertinggal	26	6735	26	6735

Lampiran 7. (Lanjutan)

λ	Partisi Data	Kelas Sebenarnya	Kelas Prediksi <i>TR-IRLS</i>		Kelas Prediksi <i>RE-WLR</i>	
			Tertinggal	Tidak tertinggal	Tertinggal	Tidak tertinggal
4	80:20	Tertinggal	64	104	65	103
		Tidak tertinggal	15	5994	15	5994
	90:10	Tertinggal	76	112	76	112
		Tidak tertinggal	18	6743	20	6741
5	50:50	Tertinggal	34	67	27	74
		Tidak tertinggal	5	3754	6	3753
	60:40	Tertinggal	39	103	38	104
		Tidak tertinggal	13	4478	13	4478
	70:30	Tertinggal	56	97	54	99
		Tidak tertinggal	16	5236	15	5237
	80:20	Tertinggal	62	106	62	106
		Tidak tertinggal	14	5995	14	5995
	90:10	Tertinggal	70	118	70	118
		Tidak tertinggal	18	6743	18	6743
6	50:50	Tertinggal	31	70	31	70
		Tidak tertinggal	5	3754	5	3754
	60:40	Tertinggal	40	102	37	105
		Tidak tertinggal	13	4478	13	4478
	70:30	Tertinggal	54	99	41	112
		Tidak tertinggal	13	5239	8	5244
	80:20	Tertinggal	47	121	59	109
		Tidak tertinggal	10	5999	13	5996
	90:10	Tertinggal	64	124	65	123
		Tidak tertinggal	17	6744	17	6744
7	50:50	Tertinggal	27	74	30	71
		Tidak tertinggal	5	3754	5	3754
	60:40	Tertinggal	38	104	35	107
		Tidak tertinggal	13	4478	11	4480
	70:30	Tertinggal	50	103	48	105
		Tidak tertinggal	10	5242	10	5242
	80:20	Tertinggal	55	113	54	114
		Tidak tertinggal	11	5998	12	5997
	90:10	Tertinggal	50	138	51	137
		Tidak tertinggal	13	6748	13	6748

Lampiran 7. (Lanjutan)

λ	Partisi Data	Kelas Sebenarnya	Kelas Prediksi <i>TR-IRLS</i>		Kelas Prediksi <i>RE-WLR</i>	
			Tertinggal	Tidak tertinggal	Tertinggal	Tidak tertinggal
8	50:50	Tertinggal	22	79	24	77
		Tidak tertinggal	5	3754	5	3754
	60:40	Tertinggal	35	107	30	112
		Tidak tertinggal	12	4479	11	4480
	70:30	Tertinggal	42	111	40	113
		Tidak tertinggal	8	5244	7	5245
	80:20	Tertinggal	47	121	47	121
		Tidak tertinggal	9	6000	9	6000
9	50:50	Tertinggal	50	138	52	136
		Tidak tertinggal	14	6747	14	6747
	60:40	Tertinggal	20	81	22	79
		Tidak tertinggal	5	3754	5	3754
	70:30	Tertinggal	32	110	29	113
		Tidak tertinggal	11	4480	10	4481
	80:20	Tertinggal	40	113	39	114
		Tidak tertinggal	7	5245	7	5245
10	50:50	Tertinggal	45	123	45	123
		Tidak tertinggal	9	6000	9	6000
	60:40	Tertinggal	49	139	49	139
		Tidak tertinggal	13	6748	14	6747
	70:30	Tertinggal	19	82	20	81
		Tidak tertinggal	5	3754	5	3754
	80:20	Tertinggal	31	111	26	116
		Tidak tertinggal	11	4480	10	4481
	90:10	Tertinggal	38	115	32	121
		Tidak tertinggal	7	5245	7	5245
	90:10	Tertinggal	42	126	43	125
		Tidak tertinggal	9	6000	9	6000
	90:10	Tertinggal	47	141	47	141
		Tidak tertinggal	13	6748	13	6748

Sumber: Hasil Pengolahan R

Lampiran 8. Confusion Matrix Model *TR-IRLS* dan *RE-WLR* Data Testing

λ	Partisi Data	Kelas Sebenarnya	Kelas Prediksi <i>TR-IRLS</i>		Kelas Prediksi <i>RE-WLR</i>	
			Tertinggal	Tidak tertinggal	Tertinggal	Tidak tertinggal
1	50:50	Tertinggal	43	64	46	61
		Tidak tertinggal	21	3733	23	3731
	60:40	Tertinggal	31	35	29	37
		Tidak tertinggal	15	3007	17	3005
	70:30	Tertinggal	22	33	23	32
		Tidak tertinggal	10	2251	11	2250
	80:20	Tertinggal	15	25	18	22
		Tidak tertinggal	9	1495	10	1494
2	50:50	Tertinggal	8	12	8	12
		Tidak tertinggal	3	749	3	749
	50:50	Tertinggal	41	66	42	65
		Tidak tertinggal	19	3735	19	3735
	60:40	Tertinggal	30	36	29	37
		Tidak tertinggal	14	3008	10	3012
	70:30	Tertinggal	21	34	21	34
		Tidak tertinggal	11	2250	10	2251
3	80:20	Tertinggal	16	24	16	24
		Tidak tertinggal	9	1495	9	1495
	90:10	Tertinggal	7	13	7	13
		Tidak tertinggal	2	750	2	750
	50:50	Tertinggal	28	79	38	69
		Tidak tertinggal	15	3739	19	3735
	60:40	Tertinggal	30	36	27	39
		Tidak tertinggal	10	3012	8	3014
4	70:30	Tertinggal	20	35	20	35
		Tidak tertinggal	10	2251	9	2252
	80:20	Tertinggal	13	27	13	27
		Tidak tertinggal	9	1495	9	1495
	90:10	Tertinggal	7	13	7	13
		Tidak tertinggal	2	750	2	750
	50:50	Tertinggal	27	80	26	81
		Tidak tertinggal	14	3740	14	3740
	60:40	Tertinggal	23	43	25	41
		Tidak tertinggal	6	3016	7	3015
	70:30	Tertinggal	20	35	20	35
		Tidak tertinggal	8	2253	8	2253

Lampiran 8. (Lanjutan)

λ	Partisi Data	Kelas Sebenarnya	Kelas Prediksi <i>TR-IRLS</i>		Kelas Prediksi <i>RE-WLR</i>	
			Tertinggal	Tidak tertinggal	Tertinggal	Tidak tertinggal
4	80:20	Tertinggal	13	27	13	27
		Tidak tertinggal	7	1497	7	1497
	90:10	Tertinggal	7	13	7	13
		Tidak tertinggal	1	751	1	751
5	50:50	Tertinggal	25	82	25	82
		Tidak tertinggal	10	3744	13	3741
	60:40	Tertinggal	26	40	22	44
		Tidak tertinggal	8	3014	5	3017
	70:30	Tertinggal	19	36	16	39
		Tidak tertinggal	8	2253	6	2255
	80:20	Tertinggal	8	32	11	29
		Tidak tertinggal	5	1499	5	1499
	90:10	Tertinggal	6	14	6	14
		Tidak tertinggal	1	751	1	751
6	50:50	Tertinggal	23	84	24	83
		Tidak tertinggal	9	3745	10	3744
	60:40	Tertinggal	24	42	22	44
		Tidak tertinggal	6	3016	5	3017
	70:30	Tertinggal	18	37	17	38
		Tidak tertinggal	7	2254	7	2254
	80:20	Tertinggal	11	29	11	29
		Tidak tertinggal	5	1499	5	1499
	90:10	Tertinggal	6	14	6	14
		Tidak tertinggal	0	752	0	752
7	50:50	Tertinggal	23	84	23	84
		Tidak tertinggal	7	3747	9	3745
	60:40	Tertinggal	22	44	21	45
		Tidak tertinggal	5	3017	4	3018
	70:30	Tertinggal	16	39	16	39
		Tidak tertinggal	6	2255	6	2255
	80:20	Tertinggal	9	31	9	31
		Tidak tertinggal	4	1500	5	1499
	90:10	Tertinggal	6	14	6	14
		Tidak tertinggal	0	752	0	752

Lampiran 8. (Lanjutan)

λ	Partisi Data	Kelas Sebenarnya	Kelas Prediksi <i>TR-IRLS</i>		Kelas Prediksi <i>RE-WLR</i>	
			Tertinggal	Tidak tertinggal	Tertinggal	Tidak tertinggal
8	50:50	Tertinggal	21	86	23	84
		Tidak tertinggal	6	3748	7	3747
	60:40	Tertinggal	22	44	21	45
		Tidak tertinggal	5	3017	4	3018
	70:30	Tertinggal	16	39	16	39
		Tidak tertinggal	6	2255	6	2255
	80:20	Tertinggal	8	32	8	32
		Tidak tertinggal	4	1500	4	1500
9	50:50	Tertinggal	18	89	22	85
		Tidak tertinggal	6	3748	7	3747
	60:40	Tertinggal	21	45	20	46
		Tidak tertinggal	5	3017	4	3018
	70:30	Tertinggal	16	39	15	49
		Tidak tertinggal	6	2255	6	2255
	80:20	Tertinggal	8	32	8	32
		Tidak tertinggal	4	1500	4	1500
10	50:50	Tertinggal	17	90	18	89
		Tidak tertinggal	6	3748	6	3748
	60:40	Tertinggal	21	45	15	51
		Tidak tertinggal	4	3018	4	3018
	70:30	Tertinggal	14	41	14	41
		Tidak tertinggal	6	2255	6	2255
	80:20	Tertinggal	8	32	8	32
		Tidak tertinggal	4	1500	4	1500
	90:10	Tertinggal	6	14	6	14
		Tidak tertinggal	0	752	0	752

Sumber: Hasil Pengolahan R

Lampiran 9. Accuracy, Sensitivity, Specificity, AUC dan G-mean Hasil Klasifikasi dengan TR-IRLS dan RE-WLR Data Training

λ	Partisi Data	TR-IRLS					RE-WLR				
		Acc	Sens	Spec	AUC	g-mean	Acc	Sens	Spec	AUC	g-mean
1	50:50	98,24%	48,51%	99,57%	0,7404	69,50%	98,24%	48,51%	99,57%	0,7404	69,50%
	60:40	97,91%	43,66%	99,62%	0,7164	65,95%	97,86%	42,96%	99,60%	0,7128	65,41%
	70:30	98,06%	45,10%	99,60%	0,7235	67,02%	98,11%	46,41%	99,62%	0,7301	67,99%
	80:20	98,11%	41,07%	99,70%	0,7039	63,99%	98,19%	46,43%	99,63%	0,7303	68,01%
	90:10	98,17%	46,28%	99,62%	0,7295	67,90%	98,17%	46,28%	99,62%	0,7295	67,90%
	Rata-rata	98,10%	44,92%	99,62%	0,7227	66,87%	98,11%	46,12%	99,61%	0,7286	67,76%
2	50:50	98,24%	43,56%	99,71%	0,7164	65,91%	98,19%	45,54%	99,60%	0,7257	67,35%
	60:40	97,86%	42,25%	99,62%	0,7094	64,88%	97,82%	38,73%	99,69%	0,6921	62,14%
	70:30	98,00%	43,14%	99,60%	0,7137	65,55%	98,02%	42,48%	99,64%	0,7106	65,06%
	80:20	98,11%	41,07%	99,70%	0,7039	63,99%	98,09%	41,07%	99,68%	0,7038	63,99%
	90:10	98,10%	40,96%	99,69%	0,7032	63,90%	98,06%	38,83%	99,70%	0,6927	62,22%
	Rata-rata	98,06%	42,20%	99,66%	0,7093	64,84%	98,03%	41,33%	99,66%	0,7050	64,15%
3	50:50	98,01%	29,70%	99,84%	0,6477	54,46%	98,16%	38,61%	99,76%	0,6919	62,07%
	60:40	97,76%	38,03%	99,64%	0,6884	61,56%	97,69%	33,80%	99,71%	0,6676	58,06%
	70:30	98,02%	41,18%	99,68%	0,7043	64,06%	97,95%	38,56%	99,68%	0,6912	62,00%
	80:20	98,07%	38,10%	99,75%	0,6892	61,64%	98,09%	38,69%	99,75%	0,6922	62,12%
	90:10	98,13%	40,43%	99,73%	0,7008	63,50%	98,10%	40,43%	99,70%	0,7006	63,49%
	Rata-rata	98,00%	37,49%	99,73%	0,6861	61,04%	98,00%	38,02%	99,72%	0,6887	61,55%
4	50:50	98,13%	33,66%	99,87%	0,6677	57,98%	97,93%	26,73%	99,84%	0,6329	51,66%
	60:40	97,50%	27,46%	99,71%	0,6359	52,33%	97,47%	26,76%	99,71%	0,6324	51,66%
	70:30	97,91%	36,60%	99,70%	0,6815	60,41%	97,89%	35,29%	99,71%	0,6750	59,32%
	80:20	98,06%	36,90%	99,77%	0,6834	60,68%	98,06%	36,90%	99,77%	0,6834	60,68%
	90:10	98,04%	37,23%	99,73%	0,6848	60,94%	98,04%	37,23%	99,73%	0,6848	60,94%
	Rata-rata	97,93%	34,37%	99,75%	0,6706	58,47%	97,88%	32,59%	99,75%	0,6617	56,85%
5	50:50	98,06%	30,69%	99,87%	0,6528	55,36%	98,06%	30,69%	99,87%	0,6528	55,36%
	60:40	97,52%	28,17%	99,71%	0,6394	53,00%	97,45%	26,06%	99,71%	0,6288	50,97%
	70:30	97,93%	35,29%	99,75%	0,6752	59,34%	97,78%	26,80%	99,85%	0,6332	51,73%
	80:20	97,88%	27,98%	99,83%	0,6390	52,85%	98,02%	35,12%	99,78%	0,6745	59,20%
	90:10	97,97%	34,04%	99,75%	0,6690	58,27%	97,99%	34,57%	99,75%	0,6716	58,73%
	Rata-rata	97,87%	31,23%	99,78%	0,6551	55,76%	97,86%	30,65%	99,79%	0,6522	55,20%
6	50:50	97,95%	26,73%	99,87%	0,6330	51,67%	98,03%	29,70%	99,87%	0,6478	54,46%
	60:40	97,47%	26,76%	99,71%	0,6324	51,66%	97,45%	24,65%	99,76%	0,6220	49,59%
	70:30	97,91%	32,68%	99,81%	0,6624	57,11%	97,87%	31,37%	99,81%	0,6559	55,96%
	80:20	97,99%	32,74%	99,82%	0,6628	57,16%	97,96%	32,14%	99,80%	0,6597	56,64%
	90:10	97,83%	26,60%	99,81%	0,6320	51,52%	97,84%	27,13%	99,81%	0,6347	52,03%
	Rata-rata	97,83%	29,10%	99,80%	0,6445	53,82%	97,83%	29,00%	99,81%	0,6440	53,74%

Lampiran 9. (Lanjutan)

λ	Partisi Data	<i>TR-IRLS</i>					<i>RE-WLR</i>				
		Acc	Sens	Spec	AUC	g-mean	Acc	Sens	Spec	AUC	g-mean
7	50:50	97,88%	23,76%	99,87%	0,6181	48,71%	97,93%	25,74%	99,87%	0,6280	50.70%
	60:40	97,47%	26,06%	99,73%	0,6289	50,98%	97,41%	23,24%	99,76%	0,6150	48.15%
	70:30	97,84%	30,07%	99,81%	0,6494	54,78%	97,82%	28,76%	99,83%	0,6429	53.58%
	80:20	97,96%	30,95%	99,83%	0,6539	55,59%	97,96%	31,55%	99,82%	0,6568	56.12%
	90:10	97,88%	29,26%	99,79%	0,6452	54,03%	97,88%	29,26%	99,79%	0,6452	54.03%
	Rata-rata	97,81%	28,02%	99,81%	0,6391	52,82%	97,80%	27,71%	99,81%	0,6376	52.52%
8	50:50	97,82%	21,78%	99,87%	0,6082	46,64%	97,88%	23,76%	99,87%	0,6181	48.71%
	60:40	97,43%	24,65%	99,73%	0,6219	49,58%	97,35%	21,13%	99,76%	0,6044	45.91%
	70:30	97,80%	27,45%	99,85%	0,6365	52,35%	97,78%	26,14%	99,87%	0,6301	51.10%
	80:20	97,90%	27,98%	99,85%	0,6391	52,85%	97,90%	27,98%	99,85%	0,6391	52.85%
	90:10	97,81%	26,60%	99,79%	0,6319	51,52%	97,84%	27,66%	99,79%	0,6373	52.54%
	Rata-rata	97,75%	25,69%	99,82%	0,6275	50,59%	97,75%	25,33%	99,83%	0,6258	50.22%
9	50:50	97,77%	19,80%	99,87%	0,5983	44,47%	97,82%	21,78%	99,87%	0,6082	46.64%
	60:40	97,39%	22,54%	99,76%	0,6115	47,41%	97,35%	20,42%	99,78%	0,6010	45.14%
	70:30	97,78%	26,14%	99,87%	0,6301	51,10%	97,76%	25,49%	99,87%	0,6268	50.45%
	80:20	97,86%	26,79%	99,85%	0,6332	51,72%	97,86%	26,79%	99,85%	0,6332	51.72%
	90:10	97,81%	26,06%	99,81%	0,6294	51,00%	97,80%	26,06%	99,79%	0,6293	51.00%
	Rata-rata	97,72%	24,27%	99,83%	0,6205	49,14%	97,72%	24,11%	99,83%	0,6197	48.99%
10	50:50	97,75%	18,81%	99,87%	0,5934	43,34%	97,77%	19,80%	99,87%	0,5983	44.47%
	60:40	97,37%	21,83%	99,76%	0,6079	46,67%	97,28%	18,31%	99,78%	0,5904	42.74%
	70:30	97,74%	24,84%	99,87%	0,6235	49,80%	97,63%	20,92%	99,87%	0,6039	45.70%
	80:20	97,81%	25,00%	99,85%	0,6243	49,96%	97,83%	25,60%	99,85%	0,6272	50.55%
	90:10	97,78%	25,00%	99,81%	0,6240	49,95%	97,78%	25,00%	99,81%	0,6240	49.95%
	Rata-rata	97,69%	23,10%	99,83%	0,6146	47,95%	97,66%	21,92%	99,83%	0,6088	46.68%
Rata-rata total		97,88%	32,04%	99,76%	0,6590	56,13%	97,86%	31,68%	99,76%	0,6572	55,77%
Median		97,89%	29,88%	99,76%	0,6485	54,62%	97,88%	29,48%	99,78%	0,6465	54,25%

Sumber: Hasil Pengolahan R

Lampiran 10. Accuracy, Sensitivity, Specificity, AUC dan G-mean Hasil Klasifikasi dengan TR-IRLS dan RE-WLR Data Testing

λ	Partisi Data	TR-IRLS					RE-WLR				
		Acc	Sens	Spec	AUC	g-mean	Acc	Sens	Spec	AUC	g-mean
1	50:50	97,80%	40,19%	99,44%	0,6981	63,22%	97,82%	42,99%	99,39%	0,7119	65,37%
	60:40	98,38%	46,97%	99,50%	0,7324	68,36%	98,25%	43,94%	99,44%	0,7169	66,10%
	70:30	98,14%	40,00%	99,56%	0,6978	63,11%	98,14%	41,82%	99,51%	0,7067	64,51%
	80:20	97,80%	37,50%	99,40%	0,6845	61,05%	97,93%	45,00%	99,34%	0,7217	66,86%
	90:10	98,06%	40,00%	99,60%	0,6980	63,12%	98,06%	40,00%	99,60%	0,6980	63,12%
	Rata-rata	98,04%	40,93%	99,50%	0,7022	63,77%	98,04%	42,75%	99,45%	0,7110	65,19%
2	50:50	97,80%	38,32%	99,49%	0,6891	61,74%	97,82%	39,25%	99,49%	0,6937	62,49%
	60:40	98,38%	45,45%	99,54%	0,7250	67,26%	98,48%	43,94%	99,67%	0,7180	66,18%
	70:30	98,06%	38,18%	99,51%	0,6885	61,64%	98,10%	38,18%	99,56%	0,6887	61,65%
	80:20	97,86%	40,00%	99,40%	0,6970	63,06%	97,86%	40,00%	99,40%	0,6970	63,06%
	90:10	98,06%	35,00%	99,73%	0,6737	59,08%	98,06%	35,00%	99,73%	0,6737	59,08%
	Rata-rata	98,03%	39,39%	99,54%	0,6946	62,56%	98,06%	39,27%	99,57%	0,6942	62,49%
3	50:50	97,57%	26,17%	99,60%	0,6288	51,05%	97,72%	35,51%	99,49%	0,6750	59,44%
	60:40	98,51%	45,45%	99,67%	0,7256	67,31%	98,48%	40,91%	99,74%	0,7032	63,88%
	70:30	98,06%	36,36%	99,56%	0,6796	60,17%	98,10%	36,36%	99,60%	0,6798	60,18%
	80:20	97,67%	32,50%	99,40%	0,6595	56,84%	97,67%	32,50%	99,40%	0,6595	56,84%
	90:10	98,06%	35,00%	99,73%	0,6737	59,08%	98,06%	35,00%	99,73%	0,6737	59,08%
	Rata-rata	97,97%	35,10%	99,59%	0,6734	58,89%	98,00%	36,06%	99,59%	0,6783	59,88%
4	50:50	97,57%	25,23%	99,63%	0,6243	50,14%	97,54%	24,30%	99,63%	0,6196	49,20%
	60:40	98,41%	34,85%	99,80%	0,6732	58,97%	98,45%	37,88%	99,77%	0,6882	61,47%
	70:30	98,14%	36,36%	99,65%	0,6800	60,20%	98,14%	36,36%	99,65%	0,6800	60,20%
	80:20	97,80%	32,50%	99,53%	0,6602	56,88%	97,80%	32,50%	99,53%	0,6602	56,88%
	90:10	98,19%	35,00%	99,87%	0,6743	59,12%	98,19%	35,00%	99,87%	0,6743	59,12%
	Rata-rata	98,02%	32,79%	99,70%	0,6624	57,06%	98,02%	33,21%	99,69%	0,6645	57,37%
5	50:50	97,62%	23,36%	99,73%	0,6155	48,27%	97,54%	23,36%	99,65%	0,6151	48,25%
	60:40	98,45%	39,39%	99,74%	0,6956	62,68%	98,41%	33,33%	99,83%	0,6658	57,69%
	70:30	98,10%	34,55%	99,65%	0,6710	58,67%	98,06%	29,09%	99,73%	0,6441	53,86%
	80:20	97,60%	20,00%	99,67%	0,5983	44,65%	97,80%	27,50%	99,67%	0,6358	52,35%
	90:10	98,06%	30,00%	99,87%	0,6493	54,74%	98,06%	30,00%	99,87%	0,6493	54,74%
	Rata-rata	97,96%	29,46%	99,73%	0,6460	53,80%	97,97%	28,66%	99,75%	0,6420	53,38%
6	50:50	97,59%	21,50%	99,76%	0,6063	46,31%	97,59%	22,43%	99,73%	0,6108	47,30%
	60:40	98,45%	36,36%	99,80%	0,6808	60,24%	98,41%	33,33%	99,83%	0,6658	57,69%
	70:30	98,10%	32,73%	99,69%	0,6621	57,12%	98,06%	30,91%	99,69%	0,6530	55,51%
	80:20	97,80%	27,50%	99,67%	0,6358	52,35%	97,80%	27,50%	99,67%	0,6358	52,35%
	90:10	98,19%	30,00%	100,00%	0,6500	54,77%	98,19%	30,00%	100,00%	0,6500	54,77%
	Rata-rata	98,02%	29,62%	99,78%	0,6470	54,16%	98,01%	28,83%	99,79%	0,6431	53,52%

Lampiran 10. (Lanjutan)

λ	Partisi Data	<i>TR-IRLS</i>					<i>RE-WLR</i>				
		Acc	Sens	Spec	AUC	g-mean	Acc	Sens	Spec	AUC	g-mean
7	50:50	97,64%	21,50%	99,81%	0,6065	46,32%	97,59%	21,50%	99,76%	0,6063	46.31%
	60:40	98,41%	33,33%	99,83%	0,6658	57,69%	98,41%	31,82%	99,87%	0,6584	56.37%
	70:30	98,06%	29,09%	99,73%	0,6441	53,86%	98,06%	29,09%	99,73%	0,6441	53.86%
	80:20	97,73%	22,50%	99,73%	0,6112	47,37%	97,67%	22,50%	99,67%	0,6108	47.36%
	90:10	98,19%	30,00%	100,00%	0,6500	54,77%	98,19%	30,00%	100,00%	0,6500	54.77%
	Rata-rata	98,01%	27,28%	99,82%	0,6355	52,00%	97,98%	26,98%	99,81%	0,6339	51.73%
8	50:50	97,62%	19,63%	99,84%	0,5973	44,27%	97,64%	21,50%	99,81%	0,6065	46.32%
	60:40	98,41%	33,33%	99,83%	0,6658	57,69%	98,41%	31,82%	99,87%	0,6584	56.37%
	70:30	98,06%	29,09%	99,73%	0,6441	53,86%	98,06%	29,09%	99,73%	0,6441	53.86%
	80:20	97,67%	20,00%	99,73%	0,5987	44,66%	97,67%	20,00%	99,73%	0,5987	44.66%
	90:10	98,19%	30,00%	100,00%	0,6500	54,77%	98,19%	30,00%	100,00%	0,6500	54.77%
	Rata-rata	97,99%	26,41%	99,83%	0,6312	51,05%	97,99%	26,48%	99,83%	0,6316	51.20%
9	50:50	97,54%	16,82%	99,84%	0,5833	40,98%	97,62%	20,56%	99,81%	0,6019	45.30%
	60:40	98,38%	31,82%	99,83%	0,6583	56,36%	98,38%	30,30%	99,87%	0,6509	55.01%
	70:30	98,06%	29,09%	99,73%	0,6441	53,86%	98,01%	27,27%	99,73%	0,6350	52.15%
	80:20	97,67%	20,00%	99,73%	0,5987	44,66%	97,67%	20,00%	99,73%	0,5987	44.66%
	90:10	98,19%	30,00%	100,00%	0,6500	54,77%	98,19%	30,00%	100,00%	0,6500	54.77%
	Rata-rata	97,97%	25,55%	99,83%	0,6269	50,13%	97,97%	25,63%	99,83%	0,6273	50.38%
10	50:50	97,51%	15,89%	99,84%	0,5786	39,83%	97,54%	16,82%	99,84%	0,5833	40.98%
	60:40	98,41%	31,82%	99,87%	0,6584	56,37%	98,22%	22,73%	99,87%	0,6130	47.64%
	70:30	97,97%	25,45%	99,73%	0,6259	50,39%	97,97%	25,45%	99,73%	0,6259	50.39%
	80:20	97,67%	20,00%	99,73%	0,5987	44,66%	97,67%	20,00%	99,73%	0,5987	44.66%
	90:10	98,19%	30,00%	100,00%	0,6500	54,77%	98,19%	30,00%	100,00%	0,6500	54.77%
	Rata-rata	97,95%	24,63%	99,84%	0,6223	49,20%	97,92%	23,00%	99,84%	0,6142	47.69%
Rata-rata total		98,00%	31,12%	99,72%	0,6542	55,26%	98,00%	31,09%	99,71%	0,6540	55,28%
Median		98,06%	31,82%	99,73%	0,6583	56,37%	98,06%	30,15%	99,73%	0,6504	54,89%

Sumber: Hasil Pengolahan R

Lampiran 11. Confusion Matrix Model *TR-IRLS* dan *RE-WLR* Data Training Stratifikasi

λ	Partisi Data	Kelas Sebenarnya	Kelas Prediksi <i>TR-IRLS</i>		Kelas Prediksi <i>RE-WLR</i>	
			Tertinggal	Tidak tertinggal	Tertinggal	Tidak tertinggal
1	50:50	Tertinggal	48	56	48	56
		Tidak tertinggal	14	3742	14	3742
	60:40	Tertinggal	57	68	57	68
		Tidak tertinggal	17	4491	17	4491
	70:30	Tertinggal	62	84	66	80
		Tidak tertinggal	15	5244	19	5240
	80:20	Tertinggal	63	103	78	88
		Tidak tertinggal	20	5990	27	5983
	90:10	Tertinggal	83	104	86	101
		Tidak tertinggal	26	6736	26	6736
2	50:50	Tertinggal	35	69	37	67
		Tidak tertinggal	11	3745	11	3745
	60:40	Tertinggal	50	75	50	75
		Tidak tertinggal	11	4497	11	4497
	70:30	Tertinggal	58	88	60	86
		Tidak tertinggal	15	5244	17	5242
	80:20	Tertinggal	61	105	75	91
		Tidak tertinggal	24	5986	24	5986
	90:10	Tertinggal	76	111	70	117
		Tidak tertinggal	23	6739	20	6742
3	50:50	Tertinggal	27	77	33	71
		Tidak tertinggal	9	3747	11	3745
	60:40	Tertinggal	47	78	48	77
		Tidak tertinggal	9	4499	9	4499
	70:30	Tertinggal	56	90	56	90
		Tidak tertinggal	15	5244	15	5244
	80:20	Tertinggal	57	109	70	96
		Tidak tertinggal	18	5992	21	5989
	90:10	Tertinggal	73	114	75	112
		Tidak tertinggal	20	6742	21	6741
4	50:50	Tertinggal	29	75	26	78
		Tidak tertinggal	10	3746	9	3747
	60:40	Tertinggal	42	83	32	93
		Tidak tertinggal	9	4499	7	4501
	70:30	Tertinggal	53	93	54	92
		Tidak tertinggal	12	5247	12	5247

Lampiran 11. (Lanjutan)

λ	Partisi Data	Kelas Sebenarnya	Kelas Prediksi <i>TR-IRLS</i>		Kelas Prediksi <i>RE-WLR</i>	
			Tertinggal	Tidak tertinggal	Tertinggal	Tidak tertinggal
4	80:20	Tertinggal	53	113	63	103
		Tidak tertinggal	16	5994	17	5993
	90:10	Tertinggal	69	118	67	120
		Tidak tertinggal	18	6744	19	6743
5	50:50	Tertinggal	26	78	27	77
		Tidak tertinggal	10	3746	10	3746
	60:40	Tertinggal	39	86	39	86
		Tidak tertinggal	8	4500	8	4500
	70:30	Tertinggal	47	99	40	106
		Tidak tertinggal	10	5249	9	5250
	80:20	Tertinggal	43	123	61	105
		Tidak tertinggal	14	5996	15	5995
	90:10	Tertinggal	64	123	64	123
		Tidak tertinggal	15	6747	16	6746
6	50:50	Tertinggal	26	78	26	78
		Tidak tertinggal	9	3747	10	3746
	60:40	Tertinggal	31	94	32	93
		Tidak tertinggal	8	4500	8	4500
	70:30	Tertinggal	44	102	44	102
		Tidak tertinggal	10	5249	10	5249
	80:20	Tertinggal	47	119	49	117
		Tidak tertinggal	14	5996	9	6001
	90:10	Tertinggal	48	139	60	127
		Tidak tertinggal	13	6749	15	6747
7	50:50	Tertinggal	25	79	25	79
		Tidak tertinggal	8	3748	8	3748
	60:40	Tertinggal	29	96	29	96
		Tidak tertinggal	8	4500	8	4500
	70:30	Tertinggal	42	104	42	104
		Tidak tertinggal	10	5249	10	5249
	80:20	Tertinggal	45	121	55	111
		Tidak tertinggal	14	5996	10	6000
	90:10	Tertinggal	59	128	46	141
		Tidak tertinggal	14	6748	12	6750

Lampiran 11. (Lanjutan)

λ	Partisi Data	Kelas Sebenarnya	Kelas Prediksi <i>TR-IRLS</i>		Kelas Prediksi <i>RE-WLR</i>	
			Tertinggal	Tidak tertinggal	Tertinggal	Tidak tertinggal
8	50:50	Tertinggal	23	81	23	81
		Tidak tertinggal	8	3748	8	3748
	60:40	Tertinggal	26	99	28	97
		Tidak tertinggal	6	4502	7	4501
	70:30	Tertinggal	38	108	39	107
		Tidak tertinggal	10	5249	10	5249
	80:20	Tertinggal	41	125	53	113
		Tidak tertinggal	14	5996	8	6002
9	50:50	Tertinggal	54	133	53	134
		Tidak tertinggal	13	6749	13	6749
	50:50	Tertinggal	21	83	21	83
		Tidak tertinggal	7	3749	8	3748
	60:40	Tertinggal	25	100	25	100
		Tidak tertinggal	6	4502	6	4502
	70:30	Tertinggal	37	109	37	109
		Tidak tertinggal	10	5249	10	5249
10	80:20	Tertinggal	37	129	52	114
		Tidak tertinggal	14	5996	8	6002
	90:10	Tertinggal	51	136	51	136
		Tidak tertinggal	13	6749	12	6750
	50:50	Tertinggal	19	85	21	83
		Tidak tertinggal	7	3749	7	3749
	60:40	Tertinggal	21	104	21	104
		Tidak tertinggal	6	4502	6	4502
	70:30	Tertinggal	35	111	36	110
		Tidak tertinggal	10	5249	10	5249
	80:20	Tertinggal	33	133	50	116
		Tidak tertinggal	14	5996	8	6002
	90:10	Tertinggal	49	138	47	140
		Tidak tertinggal	13	6749	12	6750

Sumber: Hasil Pengolahan R

Lampiran 12. *Confusion Matrix* Model *TR-IRLS* dan *RE-WLR* Data Testing Stratifikasi

λ	Partisi Data	Kelas Sebenarnya	Kelas Prediksi <i>TR-IRLS</i>		Kelas Prediksi <i>RE-WLR</i>	
			Tertinggal	Tidak tertinggal	Tertinggal	Tidak tertinggal
1	50:50	Tertinggal	47	57	47	57
		Tidak tertinggal	15	3742	15	3742
	60:40	Tertinggal	37	46	37	46
		Tidak tertinggal	16	2989	15	2990
	70:30	Tertinggal	22	40	24	38
		Tidak tertinggal	9	2245	10	2244
	80:20	Tertinggal	15	27	16	26
		Tidak tertinggal	1	1502	7	1496
2	50:50	Tertinggal	43	61	43	61
		Tidak tertinggal	14	3743	15	3742
	60:40	Tertinggal	34	49	34	49
		Tidak tertinggal	12	2993	12	2993
	70:30	Tertinggal	22	40	23	39
		Tidak tertinggal	9	2245	9	2245
	80:20	Tertinggal	15	27	13	29
		Tidak tertinggal	1	1502	6	1497
3	50:50	Tertinggal	24	80	35	69
		Tidak tertinggal	6	3751	10	3747
	60:40	Tertinggal	30	53	31	52
		Tidak tertinggal	11	2994	11	2994
	70:30	Tertinggal	20	42	20	42
		Tidak tertinggal	8	2246	8	2246
	80:20	Tertinggal	15	27	13	29
		Tidak tertinggal	1	1502	7	1496
4	50:50	Tertinggal	27	77	20	84
		Tidak tertinggal	8	3749	5	3752
	60:40	Tertinggal	30	53	24	59
		Tidak tertinggal	8	2997	9	2996
	70:30	Tertinggal	17	45	18	44
		Tidak tertinggal	8	2246	8	2246

Lampiran 12. (Lanjutan)

λ	Partisi Data	Kelas Sebenarnya	Kelas Prediksi <i>TR-IRLS</i>		Kelas Prediksi <i>RE-WLR</i>	
			Tertinggal	Tidak tertinggal	Tertinggal	Tidak tertinggal
4	80:20	Tertinggal	14	28	12	30
		Tidak tertinggal	1	1502	7	1496
	90:10	Tertinggal	7	14	6	15
		Tidak tertinggal	3	748	3	748
5	50:50	Tertinggal	23	81	25	79
		Tidak tertinggal	7	3750	7	3750
	60:40	Tertinggal	25	58	27	56
		Tidak tertinggal	8	2997	8	2997
	70:30	Tertinggal	13	49	13	49
		Tidak tertinggal	8	2246	7	2247
	80:20	Tertinggal	13	29	10	32
		Tidak tertinggal	1	1502	6	1497
	90:10	Tertinggal	6	15	6	15
		Tidak tertinggal	3	748	3	748
6	50:50	Tertinggal	22	82	22	82
		Tidak tertinggal	7	3750	7	3750
	60:40	Tertinggal	23	60	23	60
		Tidak tertinggal	8	2997	8	2997
	70:30	Tertinggal	12	50	12	50
		Tidak tertinggal	5	2249	5	2249
	80:20	Tertinggal	13	29	5	37
		Tidak tertinggal	1	1502	4	1499
	90:10	Tertinggal	6	15	6	15
		Tidak tertinggal	2	749	3	748
7	50:50	Tertinggal	18	86	7	86
		Tidak tertinggal	7	3750	18	3750
	60:40	Tertinggal	22	61	23	60
		Tidak tertinggal	8	2997	8	2997
	70:30	Tertinggal	12	50	12	50
		Tidak tertinggal	5	2249	5	2249
	80:20	Tertinggal	13	29	9	33
		Tidak tertinggal	0	1503	5	1498
	90:10	Tertinggal	6	15	3	18
		Tidak tertinggal	1	750	2	749

Lampiran 12. (Lanjutan)

λ	Partisi Data	Kelas Sebenarnya	Kelas Prediksi <i>TR-IRLS</i>		Kelas Prediksi <i>RE-WLR</i>	
			Tertinggal	Tidak tertinggal	Tertinggal	Tidak tertinggal
8	50:50	Tertinggal	16	88	17	87
		Tidak tertinggal	6	3751	6	3751
	60:40	Tertinggal	21	62	21	62
		Tidak tertinggal	8	2997	8	2997
	70:30	Tertinggal	11	51	11	51
		Tidak tertinggal	5	2249	5	2249
	80:20	Tertinggal	11	31	7	35
		Tidak tertinggal	0	1503	5	1498
9	50:50	Tertinggal	15	89	15	89
		Tidak tertinggal	6	3751	6	3751
	60:40	Tertinggal	20	63	20	63
		Tidak tertinggal	8	2997	8	2997
	70:30	Tertinggal	11	51	11	51
		Tidak tertinggal	4	2250	5	2249
	80:20	Tertinggal	11	31	7	35
		Tidak tertinggal	0	1503	5	1498
10	50:50	Tertinggal	14	90	15	89
		Tidak tertinggal	5	3752	5	3752
	60:40	Tertinggal	20	63	20	63
		Tidak tertinggal	8	2997	8	2997
	70:30	Tertinggal	10	52	10	52
		Tidak tertinggal	4	2250	4	2250
	80:20	Tertinggal	11	31	5	37
		Tidak tertinggal	0	1503	5	1498
	90:10	Tertinggal	6	15	4	17
		Tidak tertinggal	1	750	3	748

Sumber: Hasil Pengolahan R

Lampiran 13. Accuracy, Sensitivity, Specificity, AUC dan G-mean Hasil Klasifikasi dengan TR-IRLS dan RE-WLR Data Training Stratifikasi

λ	Partisi Data	TR-IRLS					RE-WLR				
		Acc	Sens	Spec	AUC	g-mean	Acc	Sens	Spec	AUC	g-mean
1	50:50	98,19%	46,15%	99,63%	0,7289	67,81%	98,19%	46,15%	99,63%	0,7289	67,81%
	60:40	98,17%	45,60%	99,62%	0,7261	67,40%	98,17%	45,60%	99,62%	0,7261	67,40%
	70:30	98,17%	42,47%	99,71%	0,7109	65,07%	98,17%	45,21%	99,64%	0,7242	67,11%
	80:20	98,01%	37,95%	99,67%	0,6881	61,50%	98,14%	46,99%	99,55%	0,7327	68,39%
	90:10	98,13%	44,39%	99,62%	0,7200	66,49%	98,17%	45,99%	99,62%	0,7280	67,68%
	Rata-rata	98,13%	43,31%	99,65%	0,7148	65,66%	98,17%	45,99%	99,61%	0,7280	67,68%
2	50:50	97,93%	33,65%	99,71%	0,6668	57,93%	97,98%	35,58%	99,71%	0,6764	59,56%
	60:40	98,14%	40,00%	99,76%	0,6988	63,17%	98,14%	40,00%	99,76%	0,6988	63,17%
	70:30	98,09%	39,73%	99,71%	0,6972	62,94%	98,09%	41,10%	99,68%	0,7039	64,00%
	80:20	97,91%	36,75%	99,60%	0,6817	60,50%	98,14%	45,18%	99,60%	0,7239	67,08%
	90:10	98,07%	40,64%	99,66%	0,7015	63,64%	98,03%	37,43%	99,70%	0,6857	61,09%
	Rata-rata	98,03%	38,15%	99,69%	0,6892	61,63%	98,08%	39,86%	99,69%	0,6977	62,98%
3	50:50	97,77%	25,96%	99,76%	0,6286	50,89%	97,88%	31,73%	99,71%	0,6572	56,25%
	60:40	98,12%	37,60%	99,80%	0,6870	61,26%	98,14%	38,40%	99,80%	0,6910	61,91%
	70:30	98,06%	38,36%	99,71%	0,6904	61,84%	98,06%	38,36%	99,71%	0,6904	61,84%
	80:20	97,94%	34,34%	99,70%	0,6702	58,51%	98,11%	42,17%	99,65%	0,7091	64,82%
	90:10	98,07%	39,04%	99,70%	0,6937	62,39%	98,09%	40,11%	99,69%	0,6990	63,23%
	Rata-rata	97,99%	35,06%	99,74%	0,6740	58,98%	98,05%	38,15%	99,71%	0,6893	61,61%
4	50:50	97,80%	27,88%	99,73%	0,6381	52,74%	97,75%	25,00%	99,76%	0,6238	49,94%
	60:40	98,01%	33,60%	99,80%	0,6670	57,91%	97,84%	25,60%	99,84%	0,6272	50,56%
	70:30	98,06%	36,30%	99,77%	0,6804	60,18%	98,08%	36,99%	99,77%	0,6838	60,75%
	80:20	97,91%	31,93%	99,73%	0,6583	56,43%	98,06%	37,95%	99,72%	0,6883	61,52%
	90:10	98,04%	36,90%	99,73%	0,6832	60,66%	98,00%	35,83%	99,72%	0,6777	59,77%
	Rata-rata	97,96%	33,32%	99,75%	0,6654	57,58%	97,94%	32,27%	99,76%	0,6602	56,51%
5	50:50	97,72%	25,00%	99,73%	0,6237	49,93%	97,75%	25,96%	99,73%	0,6285	50,88%
	60:40	97,97%	31,20%	99,82%	0,6551	55,81%	97,97%	31,20%	99,82%	0,6551	55,81%
	70:30	97,98%	32,19%	99,81%	0,6600	56,68%	97,87%	27,40%	99,83%	0,6361	52,30%
	80:20	97,78%	25,90%	99,77%	0,6284	50,84%	98,06%	36,75%	99,75%	0,6825	60,54%
	90:10	98,01%	34,22%	99,78%	0,6700	58,44%	98,00%	34,22%	99,76%	0,6699	58,43%
	Rata-rata	97,89%	29,70%	99,78%	0,6474	54,34%	97,93%	31,11%	99,78%	0,6544	55,59%
6	50:50	97,75%	25,00%	99,76%	0,6238	49,94%	97,72%	25,00%	99,73%	0,6237	49,93%
	60:40	97,80%	24,80%	99,82%	0,6231	49,76%	97,82%	25,60%	99,82%	0,6271	50,55%
	70:30	97,93%	30,14%	99,81%	0,6497	54,84%	97,93%	30,14%	99,81%	0,6497	54,84%
	80:20	97,85%	28,31%	99,77%	0,6404	53,15%	97,96%	29,52%	99,85%	0,6468	54,29%
	90:10	97,81%	25,67%	99,81%	0,6274	50,62%	97,96%	32,09%	99,78%	0,6593	56,58%
	Rata-rata	97,83%	26,78%	99,79%	0,6329	51,66%	97,88%	28,47%	99,80%	0,6413	53,24%

Lampiran 13. (Lanjutan)

λ	Partisi Data	<i>TR-IRLS</i>					<i>RE-WLR</i>				
		Acc	Sens	Spec	AUC	g-mean	Acc	Sens	Spec	AUC	g-mean
7	50:50	97,75%	24,04%	99,79%	0,6191	48,98%	97,75%	24,04%	99,79%	0,6191	48,98%
	60:40	97,76%	23,20%	99,82%	0,6151	48,12%	97,76%	23,20%	99,82%	0,6151	48,12%
	70:30	97,89%	28,77%	99,81%	0,6429	53,58%	97,89%	28,77%	99,81%	0,6429	53,58%
	80:20	97,81%	27,11%	99,77%	0,6344	52,01%	98,04%	33,13%	99,83%	0,6648	57,51%
	90:10	97,96%	31,55%	99,79%	0,6567	56,11%	97,80%	24,60%	99,82%	0,6221	49,55%
	Rata-rata	97,83%	26,93%	99,80%	0,6336	51,76%	97,85%	26,75%	99,82%	0,6328	51,55%
8	50:50	97,69%	22,12%	99,79%	0,6095	46,98%	97,69%	22,12%	99,79%	0,6095	46,98%
	60:40	97,73%	20,80%	99,87%	0,6033	45,58%	97,76%	22,40%	99,84%	0,6112	47,29%
	70:30	97,82%	26,03%	99,81%	0,6292	50,97%	97,84%	26,71%	99,81%	0,6326	51,63%
	80:20	97,75%	24,70%	99,77%	0,6223	49,64%	98,04%	31,93%	99,87%	0,6590	56,47%
	90:10	97,90%	28,88%	99,81%	0,6434	53,69%	97,88%	28,34%	99,81%	0,6408	53,19%
	Rata-rata	97,78%	24,50%	99,81%	0,6216	49,37%	97,84%	26,30%	99,82%	0,6306	51,11%
9	50:50	97,67%	20,19%	99,81%	0,6000	44,89%	97,64%	20,19%	99,79%	0,5999	44,89%
	60:40	97,71%	20,00%	99,87%	0,5993	44,69%	97,71%	20,00%	99,87%	0,5993	44,69%
	70:30	97,80%	25,34%	99,81%	0,6258	50,29%	97,80%	25,34%	99,81%	0,6258	50,29%
	80:20	97,68%	22,29%	99,77%	0,6103	47,16%	98,02%	31,33%	99,87%	0,6560	55,93%
	90:10	97,86%	27,27%	99,81%	0,6354	52,17%	97,87%	27,27%	99,82%	0,6355	52,18%
	Rata-rata	97,74%	23,02%	99,81%	0,6142	47,84%	97,81%	24,83%	99,83%	0,6233	49,60%
10	50:50	97,62%	18,27%	99,81%	0,5904	42,70%	97,67%	20,19%	99,81%	0,6000	44,89%
	60:40	97,63%	16,80%	99,87%	0,5833	40,96%	97,63%	16,80%	99,87%	0,5833	40,96%
	70:30	97,76%	23,97%	99,81%	0,6189	48,92%	97,78%	24,66%	99,81%	0,6223	49,61%
	80:20	97,62%	19,88%	99,77%	0,5982	44,53%	97,99%	30,12%	99,87%	0,6499	54,85%
	90:10	97,83%	26,20%	99,81%	0,6301	51,14%	97,81%	25,13%	99,82%	0,6248	50,09%
	Rata-rata	97,69%	21,02%	99,81%	0,6042	45,65%	97,78%	23,38%	99,84%	0,6161	48,08%
Rata-rata total		97,89%	30,18%	99,76%	0,6497	54,45%	97,93%	31,71%	99,77%	0,6574	55,79%
Median		97,87%	28,54%	99,77%	0,6416	53,37%	97,96%	30,67%	99,79%	0,6525	55,33%

Sumber: Hasil Pengolahan R

Lampiran 14. Accuracy, Sensitivity, Specificity, AUC dan G-mean Hasil Klasifikasi dengan TR-IRLS dan RE-WLR Data Testing Stratifikasi

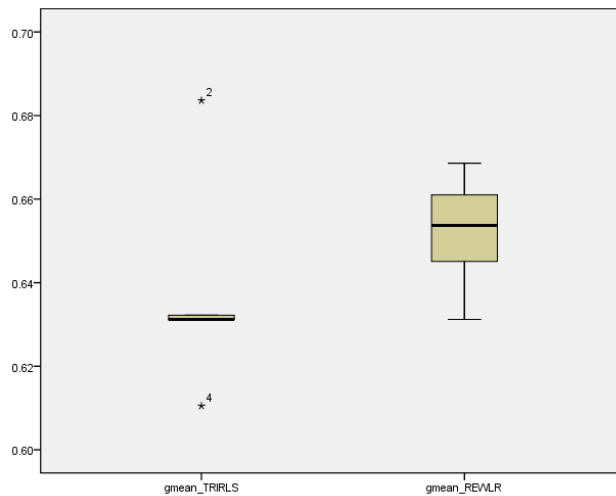
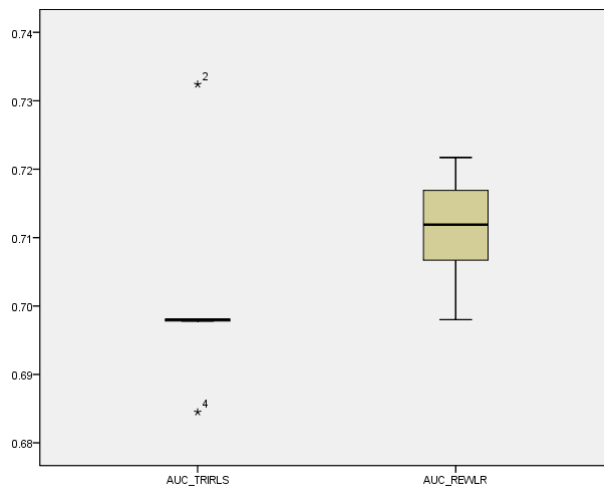
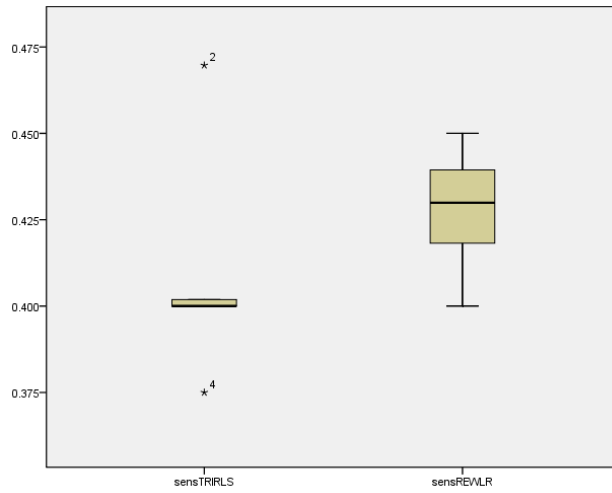
λ	Partisi Data	TR-IRLS					RE-WLR				
		Acc	Sens	Spec	AUC	g-mean	Acc	Sens	Spec	AUC	g-mean
1	50:50	98,14%	45,19%	99,60%	0,7240	67,09%	98,14%	45,19%	99,60%	0,7240	67,09%
	60:40	97,99%	44,58%	99,47%	0,7202	66,59%	98,02%	44,58%	99,50%	0,7204	66,60%
	70:30	97,88%	35,48%	99,60%	0,6754	59,45%	97,93%	38,71%	99,56%	0,6913	62,08%
	80:20	98,19%	35,71%	99,93%	0,6782	59,74%	97,86%	38,10%	99,53%	0,6881	61,58%
	90:10	97,80%	38,10%	99,47%	0,6878	61,56%	97,93%	38,10%	99,60%	0,6885	61,60%
	Rata-rata	98,00%	39,81%	99,61%	0,6971	62,89%	97,98%	40,93%	99,56%	0,7025	63,79%
2	50:50	98,06%	41,35%	99,63%	0,7049	64,18%	98,03%	41,35%	99,60%	0,7047	64,17%
	60:40	98,02%	40,96%	99,60%	0,7028	63,88%	98,02%	40,96%	99,60%	0,7028	63,88%
	70:30	97,88%	35,48%	99,60%	0,6754	59,45%	97,93%	37,10%	99,60%	0,6835	60,79%
	80:20	98,19%	35,71%	99,93%	0,6782	59,74%	97,73%	30,95%	99,60%	0,6528	55,52%
	90:10	97,80%	33,33%	99,60%	0,6647	57,62%	97,67%	28,57%	99,60%	0,6409	53,35%
	Rata-rata	97,99%	37,37%	99,67%	0,6852	60,97%	97,88%	35,79%	99,60%	0,6769	59,54%
3	50:50	97,77%	23,08%	99,84%	0,6146	48,00%	97,95%	33,65%	99,73%	0,6669	57,93%
	60:40	97,93%	36,14%	99,63%	0,6789	60,01%	97,96%	37,35%	99,63%	0,6849	61,00%
	70:30	97,84%	32,26%	99,65%	0,6595	56,70%	97,84%	32,26%	99,65%	0,6595	56,70%
	80:20	98,19%	35,71%	99,93%	0,6782	59,74%	97,67%	30,95%	99,53%	0,6524	55,51%
	90:10	97,80%	33,33%	99,60%	0,6647	57,62%	97,93%	38,10%	99,60%	0,6885	61,60%
	Rata-rata	97,91%	32,11%	99,73%	0,6592	56,41%	97,87%	34,46%	99,63%	0,6705	58,55%
4	50:50	97,80%	25,96%	99,79%	0,6287	50,90%	97,69%	19,23%	99,87%	0,5955	43,82%
	60:40	98,02%	36,14%	99,73%	0,6794	60,04%	97,80%	28,92%	99,70%	0,6431	53,69%
	70:30	97,71%	27,42%	99,65%	0,6353	52,27%	97,75%	29,03%	99,65%	0,6434	53,79%
	80:20	98,12%	33,33%	99,93%	0,6663	57,72%	97,61%	28,57%	99,53%	0,6405	53,33%
	90:10	97,80%	33,33%	99,60%	0,6647	57,62%	97,67%	28,57%	99,60%	0,6409	53,35%
	Rata-rata	97,89%	31,24%	99,74%	0,6549	55,71%	97,70%	26,86%	99,67%	0,6327	51,60%
5	50:50	97,72%	22,12%	99,81%	0,6096	46,98%	97,77%	24,04%	99,81%	0,6193	48,98%
	60:40	97,86%	30,12%	99,73%	0,6493	54,81%	97,93%	32,53%	99,73%	0,6613	56,96%
	70:30	97,54%	20,97%	99,65%	0,6031	45,71%	97,58%	20,97%	99,69%	0,6033	45,72%
	80:20	98,06%	30,95%	99,93%	0,6544	55,62%	97,54%	23,81%	99,60%	0,6171	48,70%
	90:10	97,67%	28,57%	99,60%	0,6409	53,35%	97,67%	28,57%	99,60%	0,6409	53,35%
	Rata-rata	97,77%	26,55%	99,75%	0,6315	51,29%	97,70%	25,98%	99,69%	0,6284	50,74%
6	50:50	97,69%	21,15%	99,81%	0,6048	45,95%	97,69%	21,15%	99,81%	0,6048	45,95%
	60:40	97,80%	27,71%	99,73%	0,6372	52,57%	97,80%	27,71%	99,73%	0,6372	52,57%
	70:30	97,63%	19,35%	99,78%	0,5957	43,95%	97,63%	19,35%	99,78%	0,5957	43,95%
	80:20	98,06%	30,95%	99,93%	0,6544	55,62%	97,35%	11,90%	99,73%	0,5582	34,46%
	90:10	97,80%	28,57%	99,73%	0,6415	53,38%	97,67%	28,57%	99,60%	0,6409	53,35%
	Rata-rata	97,79%	25,55%	99,80%	0,6267	50,29%	97,63%	21,74%	99,73%	0,6074	46,05%

Lampiran 14. (Lanjutan)

λ	Partisi Data	<i>TR-IRLS</i>					<i>RE-WLR</i>				
		Acc	Sens	Spec	AUC	g-mean	Acc	Sens	Spec	AUC	g-mean
7	50:50	97,59%	17,31%	99,81%	0,5856	41,56%	97,59%	17,31%	99,81%	0,5856	41,56%
	60:40	97,77%	26,51%	99,73%	0,6312	51,42%	97,80%	27,71%	99,73%	0,6372	52,57%
	70:30	97,63%	19,35%	99,78%	0,5957	43,95%	97,63%	19,35%	99,78%	0,5957	43,95%
	80:20	98,12%	30,95%	100,00%	0,6548	55,63%	97,54%	21,43%	99,67%	0,6055	46,21%
	90:10	97,93%	28,57%	99,87%	0,6422	53,42%	97,41%	14,29%	99,73%	0,5701	37,75%
	Rata-rata	97,81%	24,54%	99,84%	0,6219	49,20%	97,59%	20,02%	99,75%	0,5988	44,41%
8	50:50	97,57%	15,38%	99,84%	0,5761	39,19%	97,59%	16,35%	99,84%	0,5809	40,40%
	60:40	97,73%	25,30%	99,73%	0,6252	50,23%	97,73%	25,30%	99,73%	0,6252	50,23%
	70:30	97,58%	17,74%	99,78%	0,5876	42,07%	97,58%	17,74%	99,78%	0,5876	42,07%
	80:20	97,99%	26,19%	100,00%	0,6310	51,18%	97,41%	16,67%	99,67%	0,5817	40,76%
	90:10	97,93%	28,57%	99,87%	0,6422	53,42%	97,54%	23,81%	99,60%	0,6171	48,70%
	Rata-rata	97,76%	22,64%	99,84%	0,6124	47,22%	97,57%	19,97%	99,72%	0,5985	44,43%
9	50:50	97,54%	14,42%	99,84%	0,5713	37,95%	97,54%	14,42%	99,84%	0,5713	37,95%
	60:40	97,70%	24,10%	99,73%	0,6192	49,02%	97,70%	24,10%	99,73%	0,6192	49,02%
	70:30	97,63%	17,74%	99,82%	0,5878	42,08%	97,58%	17,74%	99,78%	0,5876	42,07%
	80:20	97,99%	26,19%	100,00%	0,6310	51,18%	97,41%	16,67%	99,67%	0,5817	40,76%
	90:10	97,93%	28,57%	99,87%	0,6422	53,42%	97,54%	23,81%	99,60%	0,6171	48,70%
	Rata-rata	97,76%	22,20%	99,85%	0,6103	46,73%	97,55%	19,35%	99,72%	0,5954	43,70%
10	50:50	97,54%	13,46%	99,87%	0,5666	36,67%	97,57%	14,42%	99,87%	0,5715	37,95%
	60:40	97,70%	24,10%	99,73%	0,6192	49,02%	97,70%	24,10%	99,73%	0,6192	49,02%
	70:30	97,58%	16,13%	99,82%	0,5798	40,13%	97,58%	16,13%	99,82%	0,5798	40,13%
	80:20	97,99%	26,19%	100,00%	0,6310	51,18%	97,28%	11,90%	99,67%	0,5579	34,45%
	90:10	97,93%	28,57%	99,87%	0,6422	53,42%	97,41%	19,05%	99,60%	0,5932	43,56%
	Rata-rata	97,75%	21,69%	99,86%	0,6077	46,08%	97,51%	17,12%	99,74%	0,5843	41,02%
Rata-rata total		97,84%	28,37%	99,77%	0,6407	52,68%	97,70%	26,22%	99,68%	0,6295	50,38%
Median		97,80%	28,57%	99,78%	0,6419	53,40%	97,67%	24,70%	99,67%	0,6222	49,63%

Sumber: Hasil Pengolahan R

Lampiran 15. Perbandingan Box Plot Ukuran Ketepatan Klasifikasi *TR-IRLS* dan *RE-WLR* pada λ Optimal (Partisi Data Random)



Lampiran 16. Uji Wilcoxon dengan SPSS (Partisi Data Random)

Descriptive Statistics

	N	Mean	Std. Deviation	Minimum	Maximum
accTRIRLS	5	.980360	.0024552	.9780	.9838
sensTRIRLS	5	.409320	.0355401	.3750	.4697
specTRIRLS	5	.995000	.0008246	.9940	.9960
AUC_TRIRLS	5	.702160	.0178825	.6845	.7324
gmean_TRIRLS	5	.637720	.0272154	.6105	.6836
accREWLR	5	.980400	.0016956	.9782	.9825
sensREWLR	5	.427500	.0193414	.4000	.4500
specREWLR	5	.994560	.0010213	.9934	.9960
AUC_REWLR	5	.711040	.0091868	.6980	.7217
gmean_REWLR	5	.651920	.0144885	.6312	.6686

Ranks

		N	Mean Rank	Sum of Ranks
accREWLR - accTRIRLS	Negative Ranks	1 ^a	2.50	2.50
	Positive Ranks	2 ^b	1.75	3.50
	Ties	2 ^c		
	Total	5		
sensREWLR - sensTRIRLS	Negative Ranks	1 ^d	3.00	3.00
	Positive Ranks	3 ^e	2.33	7.00
	Ties	1 ^f		
	Total	5		
specREWLR - specTRIRLS	Negative Ranks	4 ^g	2.50	10.00
	Positive Ranks	0 ^h	.00	.00
	Ties	1 ⁱ		
	Total	5		
AUC_REWLR - AUC_TRIRLS	Negative Ranks	1 ^j	3.00	3.00
	Positive Ranks	3 ^k	2.33	7.00
	Ties	1 ^l		
	Total	5		
gmean_REWLR - gmean_TRIRLS	Negative Ranks	1 ^m	3.00	3.00
	Positive Ranks	3 ⁿ	2.33	7.00
	Ties	1 ^o		
	Total	5		

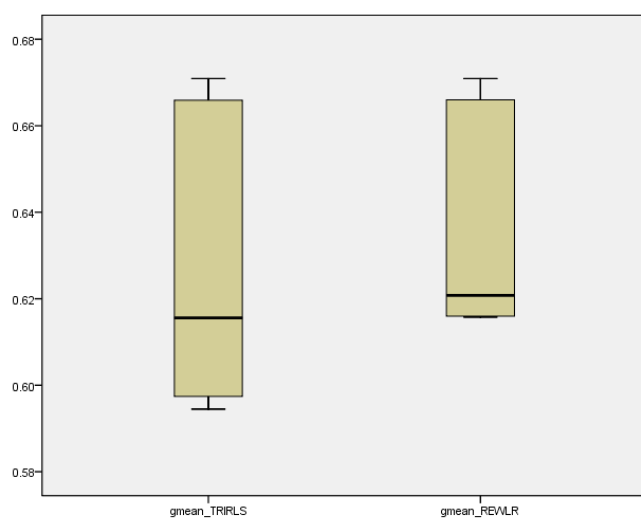
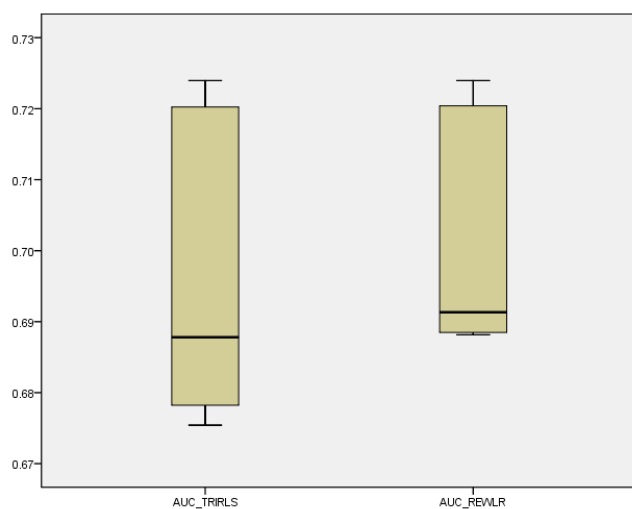
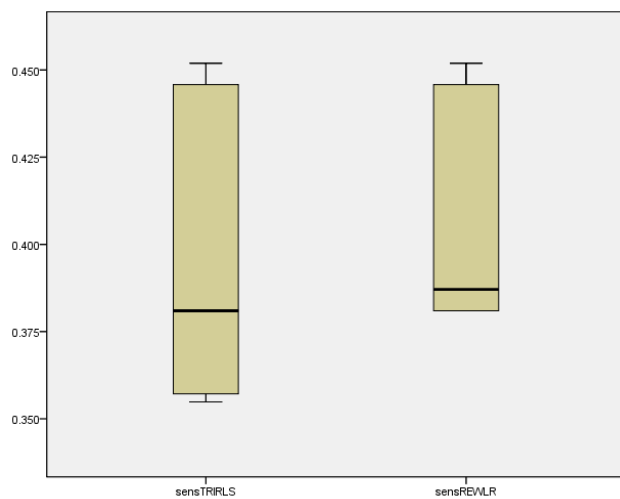
- a. accREWLR < accTRIRLS
 b. accREWLR > accTRIRLS
 c. accREWLR = accTRIRLS
 d. sensREWLR < sensTRIRLS
 e. sensREWLR > sensTRIRLS
 f. sensREWLR = sensTRIRLS
 g. specREWLR < specTRIRLS
 h. specREWLR > specTRIRLS
 i. specREWLR = specTRIRLS
 j. AUC_REWLR < AUC_TRIRLS
 k. AUC_REWLR > AUC_TRIRLS
 l. AUC_REWLR = AUC_TRIRLS
 m. gmean_REWLR < gmean_TRIRLS
 n. gmean_REWLR > gmean_TRIRLS
 o. gmean_REWLR = gmean_TRIRLS

Test Statistics^c

	accREWLR - accTRIRLS	sensREWLR - sensTRIRLS	specREWLR - specTRIRLS	AUC_REWLR - AUC_TRIRLS	gmean_REWLR - gmean_TRIRLS
Z	-.272 ^a	-.730 ^a	-1.857 ^b	-.730 ^a	-.730 ^a
Asymp. Sig. (2-tailed)	.785	.465	.063	.465	.465

- a. Based on negative ranks.
 b. Based on positive ranks.
 c. Wilcoxon Signed Ranks Test

Lampiran 17. Perbandingan Box Plot Ukuran Ketepatan Klasifikasi *TR-IRLS* dan *RE-WLR* pada λ optimal (Partisi Data Stratifikasi)



Lampiran 18. Uji Wilcoxon dengan SPSS (Partisi Data Stratifikasi)

Descriptive Statistics

	N	Mean	Std. Deviation	Minimum	Maximum
accTRIRLS	5	.979995	.0016416	.9780	.9819
sensTRIRLS	5	.398128	.0474702	.3548	.4519
specTRIRLS	5	.996140	.0019063	.9947	.9993
AUC_TRIRLS	5	.697134	.0232841	.6754	.7240
gmean_TRIRLS	5	.628855	.0370333	.5945	.6709
accREWLR	5	.979758	.0010600	.9786	.9814
sensREWLR	5	.409342	.0362212	.3810	.4519
specREWLR	5	.995585	.0004321	.9950	.9960
AUC_REWLR	5	.702464	.0180856	.6881	.7240
gmean_REWLR	5	.637891	.0280273	.6158	.6709

Ranks

		N	Mean Rank	Sum of Ranks
accREWLR - accTRIRLS	Negative Ranks	1 ^a	4.00	4.00
	Positive Ranks	3 ^b	2.00	6.00
	Ties	1 ^c		
	Total	5		
sensREWLR - sensTRIRLS	Negative Ranks	0 ^d	.00	.00
	Positive Ranks	2 ^e	1.50	3.00
	Ties	3 ^f		
	Total	5		
specREWLR - specTRIRLS	Negative Ranks	2 ^g	3.00	6.00
	Positive Ranks	2 ^h	2.00	4.00
	Ties	1 ⁱ		
	Total	5		
AUC_REWLR - AUC_TRIRLS	Negative Ranks	0 ^j	.00	.00
	Positive Ranks	4 ^k	2.50	10.00
	Ties	1 ^l		
	Total	5		
gmean_REWLR - gmean_TRIRLS	Negative Ranks	0 ^m	.00	.00
	Positive Ranks	4 ⁿ	2.50	10.00
	Ties	1 ^o		
	Total	5		

- a. accREWLR < accTRIRLS
 b. accREWLR > accTRIRLS
 c. accREWLR = accTRIRLS
 d. sensREWLR < sensTRIRLS
 e. sensREWLR > sensTRIRLS
 f. sensREWLR = sensTRIRLS
 g. specREWLR < specTRIRLS
 h. specREWLR > specTRIRLS
 i. specREWLR = specTRIRLS
 j. AUC_REWLR < AUC_TRIRLS
 k. AUC_REWLR > AUC_TRIRLS
 l. AUC_REWLR = AUC_TRIRLS
 m. gmean_REWLR < gmean_TRIRLS
 n. gmean_REWLR > gmean_TRIRLS
 o. gmean_REWLR = gmean_TRIRLS

Test Statistics^c

	accREWLR - accTRIRLS	sensREWLR - sensTRIRLS	specREWLR - specTRIRLS	AUC_REWLR - AUC_TRIRLS	gmean_REWLR - gmean_TRIRLS
Z	-.365 ^a	-1.342 ^a	-.365 ^b	-1.826 ^a	-1.826 ^a
Asymp. Sig. (2-tailed)	.715	.180	.715	.068	.068

- a. Based on negative ranks.
 b. Based on positive ranks.
 c. Wilcoxon Signed Ranks Test

Lampiran 19. Urutan Signifikansi Variabel Berdasarkan Nilai $\tilde{\beta}_j$

No	Nilai	Var	Keterangan
1	2.89152787	X7	Ketersediaan Puskesmas
2	1.8285665	X38	Kelengkapan Pemerintahan Desa
3	1.71287467	X20	Bahan Bakar Untuk Memasak
4	1.37022066	X27	Aksesibilitas Jalan
5	1.36875489	X24	Kualitas Fasilitas Komunikasi Seluler
6	1.34128968	X19	Kondisi Penerangan di Jalan Utama
7	1.28317778	X31	Biaya per Kilometer Transportasi ke Kantor Camat
8	1.26837481	X4	Ketersediaan SMA Sederajat
9	1.19708282	X21	Sumber Air Untuk Minum
10	1.1808116	X1	Ketersediaan TK/RA/BA
11	1.16240207	X42	Kualitas SDM Sekretaris Desa
12	1.08658416	X39	Otonomi Desa
13	1.05706836	X28	Ketersediaan Angkutan Umum
14	1.02636249	X25	Ketersediaan Fasilitas Internet dan Pengiriman Pos atau Barang
15	-1.02434392	X30	Waktu Tempuh per Kilometer Transportasi ke Kantor Camat
16	0.84983308	X9	Ketersediaan Tempat Praktek Dokter
17	0.80920328	X26	Lalu Lintas dan Kualitas Jalan
18	-0.72527261	X32	Waktu Tempuh per Kilometer Transportasi ke Kantor Bupati/Walikota
19	0.72525118	X15	Ketersediaan Restoran, Rumah Makan atau Warung Makan
20	0.64811814	X3	Ketersediaan SMP Sederajat
21	0.62332951	X36	Ketersediaan Fasilitas Olah Raga
22	0.6046837	X8	Ketersediaan Poliklinik/Balai Pengobatan
23	0.59630616	X23	Fasilitas Buang Air Besar
24	0.50319082	X10	Ketersediaan Tempat Praktek Bidan
25	0.44817788	X37	Keberadaan Kelompok Kegiatan Olah Raga
26	0.36744685	X40	Aset/Kekayaan Desa
27	0.34089745	X18	Elektrifikasi
28	0.28964751	X17	Ketersediaan Bank
29	0.28958918	X29	Operasional Angkutan Umum
30	0.22080506	X11	Ketersediaan Poskesdes atau Polindes
31	0.17774125	X34	Penanganan Kejadian Luar Biasa (KLB)
32	0.15422046	X41	Kualitas SDM Kepala Desa
33	0.14208049	X6	Ketersediaan Rumah Sakit Bersalin
34	0.12994593	X5	Ketersediaan Rumah Sakit
35	-0.10711376	X2	Ketersediaan SD Sederajat
36	0.10166546	X22	Sumber Air Untuk Mandi/Cuci
37	-0.08218794	X16	Ketersediaan Akomodasi Hotel atau Penginapan
38	-0.07174735	X13	Ketersediaan Pertokoan, Minimarket atau Toko Kelontong
39	0.07103133	X14	Ketersediaan Pasar
40	0.04016721	X12	Ketersediaan Apotek
41	0.03262516	X35	Penanganan Gizi Buruk
42	-0.00258227	X33	Biaya per Kilometer Transportasi ke Kantor Bupati/Walikota

DAFTAR PUSTAKA

- Agresti, A., (1996), *An Introduction to Categorical Data Analysis*, John Willey and Son, Inc., New York.
- Badan Perencanaan Pembangunan Nasional dan Badan Pusat Statistik, (2015), *Indeks Pembangunan Desa 2014 “Tantangan Pemenuhan Standar Pelayanan Minimum Desa”*, Bappenas, Jakarta.
- Badan Pusat Statistik, (2005), *Identifikasi dan Penentuan Desa Tertinggal 2002*, Badan Pusat Statistik, Jakarta.
- , (2008), *Analisis dan Penghitungan Tingkat Kemiskinan 2008*, Badan Pusat Statistik, Jakarta.
- , (2014), *Pedoman Pendataan PODES 2014*, Badan Pusat Statistik, Jakarta.
- BAPPENAS, (1993), *Panduan Pelaksanaan Program IDT 1994-1999*, Jakarta.
- Cordeiro, G.M. dan McCullagh, P. (1991), “Bias correction in generalized linear models”, *Journal of Royal Statistical Society*, 53(3):, hal. 629–643.
- Dimulyo, S. (2009), “Penggunaan Geographically Weighted Regression-Kriging untuk Klasifikasi Desa Tertinggal”, *Seminar Nasional Aplikasi Teknologi Informasi (SNATI) 2009*, Yogyakarta.
- Dirjen Pemberdayaan Masyarakat Desa (PMD) Kemendagri, (2012), *Profil Desa dan Kelurahan 2011: Data Dasar Tipologi, Klasifikasi, Kategori Desa dan Kelurahan Menurut Provinsi*, Dirjen PMD Kemendagri, Jakarta.
- Djuraidah, A. (2009), “Analisis Status Ketertinggalan Daerah dengan Analisis Diskriminan”, *Seminar Nasional Matematika dan Pendidikan Matematika, UNY*, Yogyakarta
- Edy, L. (2009), “Pencapaian Pembangunan Daerah Tertinggal Lima Tahun Terakhir”, *Jurnal Sekretariat Negara*, No : 13, Agustus 2009.
- Fawcett, T. (2006), “An Introduction to ROC Analysis”, *Pattern Recognition Letters* **27** (8), hal 861 – 874.
- Han, J dan Kamber M, (2001), *Data Mining Concepts and Techniques*, Academic

Press, United States of America.

- Hanifah, E. (2010), *Metode Latent Class Cluster Untuk Variabel Indikator Bertipe Campuran Dalam Rangka Pengelompokan Desa Tertinggal*, Tesis, Universitas Padjajaran.
- Haerdle, W.K., Prastyo, D.D., dan Hafner, C.M. (2014), "Support Vector Machines with Evolutionary Model Selection for Default Prediction", dalam *The Oxford Handbook of Applied Nonparametric and Semiparametric Econometric and Statistics*, eds: Racine, J.S., Su, L., dan Ullah, A., Oxford University Press, New York, hal. 346-373.
- Haerdle, W.K. dan Prastyo, D.D. (2014), "Embedded Predictor Selection for Default Risk Calculation: A Southeast Asian Industry Study," dalam *Handbook of Asian Finance Vol. 1: Financial Market and Sovereign Wealth Funds*, eds: Chuen, D.L.K. dan Gregoriou, G.N., Academic Press, San Diego, hal. 131-148.
- Hastie, T., Tibshirani, R. dan Friedman, J. (2001), *The Elements of Statistical Learning*. Springer Verlag.
- Hidayat, S. (2008), *Permodelan Desa Tertinggal di Jawa Barat Tahun 2005 dengan Pendekatan MARS*, Tesis, ITS, Surabaya.
- Hilas, C.S. (2009), "Designing Expert System for Fraud Detection in Private Telecommunication Networks", *Expert Systems with Applications*, Vol. 36 (9), hal. 11559-11569.
- Hosmer dan Lemeshow, (2000), *Applied Logistic Regression 2nd Edition*, John Willey and Son, Inc., New York.
- Huberty, C.J. (1989), "Problems with Stepwise Methods: Better alternatives", *Advances in Social Science Methodology*, Vol. 1, hal. 43–70.
- Karsmakers, P., Pelckmans, K. dan Suykens, J.A.K. (2007), "Multi-class kernel logistic regression: a fixed-size implementation", *International Joint Conference on Neural Networks*, hal. 1756–1761.
- King, Gary dan Zheng, Langche (2001), "Logistic Regression in Rare Events Data", *Political Analysis*, Vol. 9, hal. 137–163.

- Komarek, P. (2004), *Logistic Regression for Data Mining and High-dimensional Classification*, Ph.D. Thesis, Carnegie Mellon University, Pittsburgh.
- Komarek, P. dan Moore, A. (2005), *Making Logistic Regression a Core Data Mining Tool*, Technical Report, Carnegie Mellon University.
- Kononenko, I. (2001), “Machine learning for medical diagnosis: history, state of the art and perspective”, *Artificial Intelligence in Medicine*, Vol.23, hal.89–109.
- Kubat, M. dan Matwin, S. (1997), “Addressing the Curse of Imbalanced Training Set: One Sided Selection”, *14th International Conference on Machine Learning*, Nashville , TN, USA, hal. 179 – 186.
- Kubat, M., Matwin, S., dan Holte, R. (1998), “Machine Learning for the Detection of Oil Spills In Satellite Radar Images”, *Machine Learning*, Vol. 30, hal.195-215.
- Kutner, M.H., Nachtsheim, C.J. dan Neter, J. (2004), *Applied Linear Regression Models Fourth Edition*, The McGraw-Hill Companies, Inc, NewYork.
- Lewis, D dan Carlett, J. (1994), “Heterogeneous Uncertainly Sampling for Supervised Learning” , *Proceedings of ICML-94, 11th International Conference on Machine Learning*, Eds: Cohen, W. dan Hirsh, H., Morgan Kaufmann, San Fransisco, hal. 148–156.
- Lin, C., Weng, R.C. dan Keerthi, S.S. (2008), “Trust Region Newton Method for Large-scale Logistic Regression”, *Journal of Machine Learning Research*, Vol 9, hal. 627-650.
- Longadge, R., Dongre, S., dan Malik, L. (2013), “Class Imbalance Problem in Data Mining: Review”. *International Journal of Computer Science and Network*, Vol. 2.
- Maalouf (2009), *Robust Weighted Kernel Logistic Regression in Imbalanced and rare events data*, Disertasi, University of Oklahoma, Oklahoma.
- Maalouf dan Trafalis (2010), “Robust weighted kernel logistic regression in imbalanced and rare events data”, *Computational Statistics and Data Analysis 2011*, Vol. 55, hal. 168-183.

- Maalouf dan Siddiqi (2014), “Weighted logistic regression for large-scale imbalanced and rare events data”, *Journal of Knowledge-Based Systems*, Vol. 59, hal. 141-148.
- McCullagh, P. dan Nelder, J. (1989), *Generalized Linear Model*. Chapman and Hall/CRC.
- Menteri Negara Pembangunan Daerah Tertinggal Republik Indonesia, (2005), *Strategi Nasional Pembangunan Daerah Tertinggal*, Kementerian Negara Pembangunan Daerah Tertinggal Republik Indonesia, Jakarta.
- Mulya, A. (2005), *Analisis Penetapan Penduduk Miskin BPS Studi Kasus di Kabupaten Kupang, NTT*, Tesis, IPB, Bogor.
- Myers, R.H., Montgomery, D.C., and Vining, G.G., (2002), *Generalized Linear Models, with Applications in Engineering and the Sciences*, John Wiley and Sons Inc., New York.
- Nooraeni, R. (2015), *Metode Cluster Menggunakan Kombinasi Algoritma Cluster K-Prototype dan Algoritma Genetika untuk Data bertipe Campuran (Studi Kasus: Pengelompokan Desa berdasarkan Indikator Ketertinggalan Desa)*, Tesis, Universitas Padjajaran, Bandung.
- Rahayu, S.P. (2012), *Logistic regression methods for classification of imbalanced data*, Tesis Ph.D, University Malaysia Pahang (UMP), Pahang.
- Rahayu, S.P., Zain, J.M., Embonga, A., Juwari, dan Purnami, S.W. (2012), “Logistic regression methods with truncated newton method”, *Procedian Engineering 2012*, Vol. 50, hal. 827-836.
- Sain, H. dan Purnami, S.W. (2013), *Combine Sampling Support Vector Machine untuk Klasifikasi Data Imbalanced*, Tesis, ITS, Surabaya.
- Triola, M.F. (2012), *Elementary Statistics 11th Edition*, Pearson Education, Inc., Boston.
- Undang-Undang Nomor 6 Tahun 2014 tentang Desa.
- Wahyuningsih, D. (2009), *Analisis Karakteristik Desa Tertinggal Dengan Structural Equation Modeling studi Kasus di Provinsi Lampung*, Tesis, ITS, Surabaya.

- Wang, Xu, dan Zhou (2015), “Large Unbalanced Credit Scoring Using Lasso-Logistic Regression Ensemble”, *PLoS ONE* 10(2): e0117844.
- Witten, I. H., Frank, E., dan Hall, M. A. (2011), *Data Mining: Practical Machine Learning Tools and Technique*, Morgan Kaufmann Publishers Inc., San Francisco.
- Wu, G. dan Chang, E. (2003), “Class-Boundary Alignment for Imbalanced Dataset Learning”, *ICML 2003 Workshoop on Learning from Imbalanced Data Sets II*, Washington.
- Yuan, Ho, dan Lin (2012), “Recent Advances of Large-Scale Linear Classification”, *Proceedings of The IEEE* 100 (9), hal. 2584–2603.

Halaman ini sengaja dikosongkan

BIOGRAFI PENULIS



Penulis dilahirkan di Tuban, Jawa Timur pada tanggal 3 April 1984. Penulis merupakan putri pertama dari pasangan Danang Sudarwoko dengan Siti Rokhanah. Penulis menempuh pendidikan formal di SDN 1 Kapas (1991-1992), SDN 1 Kedungadem (1992-1994), SDN 2 Kapas (1994-1996), SLTPN 1 Kapas (1996-1999), SMUN 1 Bojonegoro (1999-2002), Komputasi Statistik di Sekolah Tinggi Ilmu Statistik (STIS) Jakarta (2003-2007). Setelah menyelesaikan pendidikan di STIS, penulis ditugaskan di BPS Kabupaten Hulu Sungai Utara dan BPS Kabupaten Tapin Kalimantan Selatan. Pada tahun 2014 penulis memperoleh kesempatan dari BPS untuk melanjutkan jenjang pendidikan S2 di Jurusan Statistika Fakultas Matematika dan Ilmu Pengetahuan Alam (FMIPA) Institut Teknologi Sepuluh Nopember (ITS) Surabaya. Pembaca yang ingin memberikan kritik, saran dan pertanyaan mengenai penelitian ini, dapat menghubungi penulis melalui email deasdian@gmail.com.